ADVENTURES IN THEORETICAL PHYSICS

Selected Papers with Commentaries

World Scientific Series in 20th Century Physics Vol. 37

Stephen L. Adler

World Scientific

ADVENTURES IN THEORETICAL PHYSICS

World Scientific Series in 20th Century Physics

Published

Vol. 20 The Origin of the Third Family
edited by O. Barnabei, L. Maiani, R. A. Ricci and F. R. Monaco

Vol. 21 Spectroscopy with Coherent Radiation — Selected Papers of Norman F. Ramsey (with Commentary)
edited by N. F. Ramsey

Vol. 22 A Quest for Symmetry — Selected Works of Bunji Sakita
edited by K. Kikkawa, M. Virasoro and S. R. Wadia

Vol. 23 Selected Papers of Kun Huang (with Commentary)
edited by B.-F. Zhu

Vol. 24 Subnuclear Physics — The First 50 Years: Highlights from Erice to ELN by A. Zichichi
edited by O. Bamabei, P. Pupillo and F. Roversi Monaco

Vol. 25 The Creation of Quantum Chromodynamics and the Effective Energy by V. N. Gribov, G. 't Hooft, G. Veneziano and V. F. Weisskopf
edited by L. N. Lipatov

Vol. 26 A Quantum Legacy — Seminal Papers of Julian Schwinger
edited by K. A. Milton

Vol. 27 Selected Papers of Richard Feynman (with Commentary)
edited by L. M. Brown

Vol. 28 The Legacy of Léon Van Hove
edited by A. Giovannini

Vol. 29 Selected Works of Emil Wolf (with Commentary)
edited by E. Wolf

Vol. 30 Selected Papers of J. Robert Schrieffer — In Celebration of His 70th Birthday
edited by N. E. Bonesteel and L. P. Gor'kov

Vol. 31 From the Preshower to the New Technologies for Supercolliders — In Honour of Antonino Zichichi
edited by B. H. Wiik, A. Wagner and H. Wenninger

Vol. 32 In Conclusion — A Collection of Summary Talks in High Energy Physics
edited by J. D. Bjorken

Vol. 33 Formation and Evolution of Black Holes in the Galaxy — Selected Papers with Commentary
edited by H. A. Bethe, G. E. Brown and C.-H. Lee

Vol. 35 A Career in Theoretical Physics, 2nd Edition
by P. W. Anderson

Vol. 36 Selected Papers (1945–1980) with Commentary
by Chen Ning Yang

Vol. 37 Adventures in Theoretical Physics — Selected Papers with Commentaries
by Stephen L. Adler

For information on Vols. 1–19, please visit http://www.worldscibooks.com/series/wsscp_series.shtml

World Scientific Series in 20th Century Physics **Vol. 37**

ADVENTURES IN THEORETICAL PHYSICS

Selected Papers with Commentaries

Stephen L. Adler

Institute for Advanced Study, Princeton

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Adler, Stephen L.

Adventures in theoretical physics : selected papers with commentaries / Stephen L. Adler.

p. cm. -- (World Scientific series in 20th century physics; v. 37)

Includes bibliographical references and index.

ISBN 981-256-370-9 -- ISBN 981-256-522-1 (pbk.)

1. Mathematical physics. 2. Physics. I. Adler, Stephen L. II. Title.

QC20.5.A35 2006

530.15--dc22

2005058116

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ADVENTURES IN THEORETICAL PHYSICS

Selected Papers with Commentaries

Printed in Singapore by B & JO Enterprise

Dedicated with love to my father Irving,
the memory of my mother Ruth,
and my sister Peggy

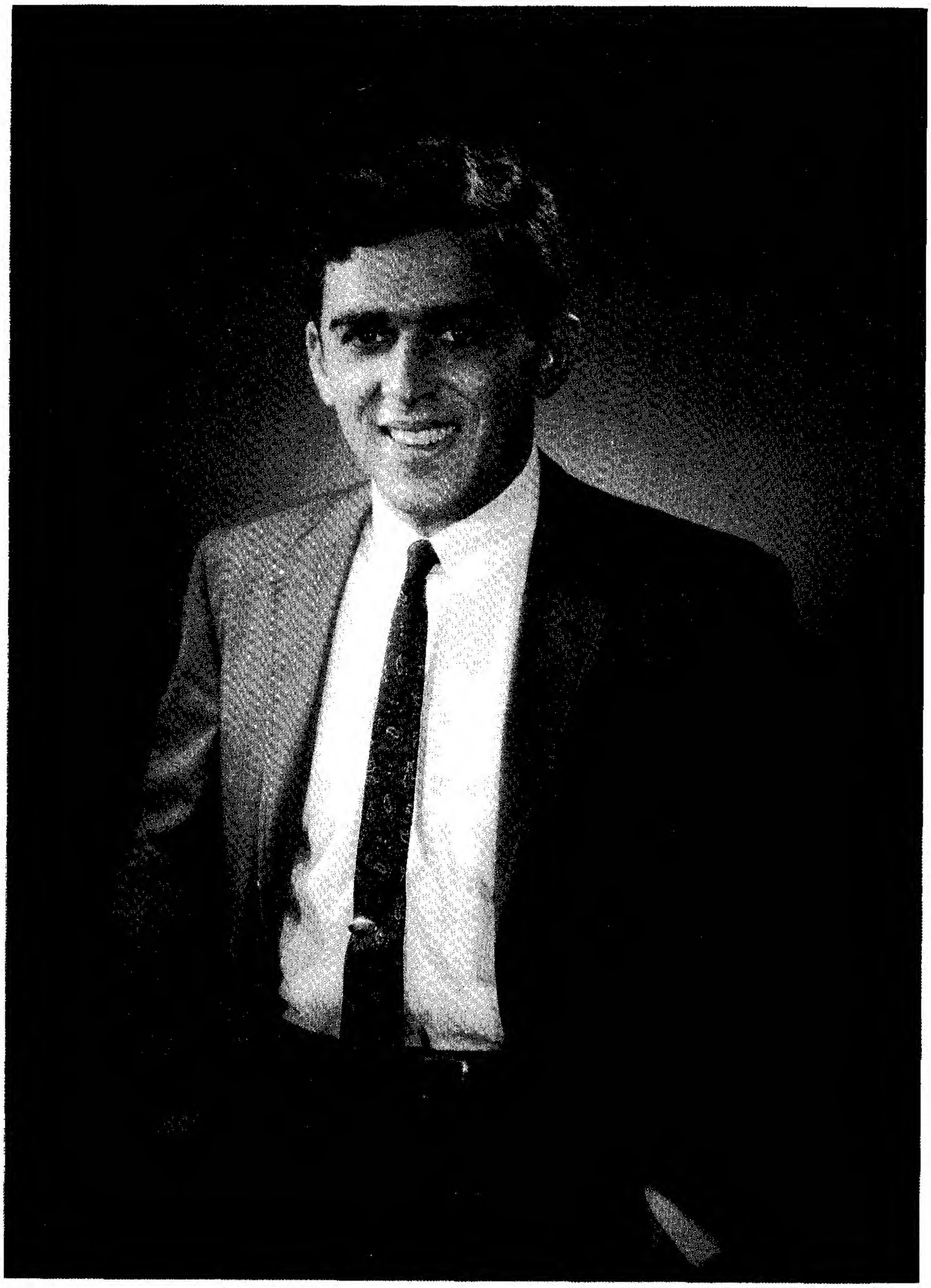


Photo by Bachrach

Stephen L. Adler, July 1966, just before moving to Princeton to join the Institute for Advanced Study. For a photo taken in 1995, see reprint R1.

Preface

When I was asked by K. K. Phua to do a book for World Scientific based on my work, he suggested a volume of essays or a reprint volume. I have decided to combine these two suggestions into one, by preparing a reprint volume with commentaries. Some of the commentaries are drawn from historical articles that I have written for publication, others are drawn from unpublished historical accounts written for institutional archives, and yet others have been written expressly for this volume. In the commentaries, I try to relate the reprinted articles to the time-line of my career, and at the same time to analyze their relations with the work of other physicists whose work influenced mine and vice versa.

In keeping with these dual aims, I have arranged the articles and the commentaries in approximately chronological order, but occasionally deviate from strict chronology in order to group topically related articles together. In choosing which articles to include, I have been guided by two generally coinciding measures, my own estimate of significance, and the citation count. However, in occasional cases I have included infrequently cited articles where I felt that there was an interesting related story to tell. Often, when finishing a line of work, I have written a long summarizing article or review; some of these are too long to be included in their entirety, and so I have included in the reprints only the sections most relevant to the narrative in the commentaries. Similarly, I have not included among the reprints the summer school lectures I have given on current algebras, anomalies, and neutrino physics, but references to them appear in the commentaries. In the last decade, I have published two books related to my work on generalized forms of quantum mechanics, and included many research results directly in these books in lieu of first writing papers. It is feasible to give only brief descriptions of these projects in the commentaries; I have included just a few papers from this period, all in the nature of follow-ons to the first book.

In both the texts of the commentaries and the reference lists that follow them, reprinted articles are identified by a sans serif R, so that for example, R1 designates the first reprinted article. Numbers in square brackets following each reference in the reference lists give the pages in the commentaries where that reference is cited. There is also an index of names following the commentaries, and a list of detailed chapter subheadings in the Table of Contents.

I wish to thank Tian-Yu Cao for a critical reading of the commentaries and much

helpful advice, Alfred Mueller for a helpful conversation on renormalon ambiguities, Richard Haymaker for a clarifying email on dual superconductivity parameters, and William Marciano, Robert Oakes, and Alberto Sirlin for calling my attention to relevant references. I also wish to thank the following people for sending me helpful comments on the initial draft of the commentaries after it was posted on the archive as hep-ph/0505177: Nikolay Achasov, Dimi Chakalov, Christopher Hill, Roman Jackiw, Andrei Kataev, Peter Minkowski, Herbert Neuberger, and Lalit Sehgal. I am grateful to Antonino Zichichi for permission to use the quote from Gilberto Bernardini in Chapter 2, to Mary Bell for permission to use the quote from John Bell in Chapter 3, to James Bjorken for permission use his quote in Chapter 3, and to Clifford Taubes for helpful email correspondence and permission to use his quotes in Chapter 7.

My editor at World Scientific, Kim Tan, has given valuable assistance throughout this project. Miriam Peterson and Margaret Best have patiently assisted in the conversion of my TeX drafts to camera-ready copy and with indexing, the latter a task that was shared with Lisa Fleischer and Michelle Sage. I am also indebted to Momota Ganguli and Judy Wilson-Smith for bibliographic searches, to Christopher McCafferty and James Stephens for help with computer problems, and to Marcia Tucker and Herman Joachim for assistance, respectively, in scanning and duplicating certain of the papers to be reprinted. Finally, I wish to express my appreciation to the Institute for Advanced Study (abbreviated throughout the commentaries as IAS) for its support of my work, first from 1966 to 1969, when I was a Long Term Member, and then from 1969 onwards, when I have been a member of the Faculty, in the School of Natural Sciences. My work has also been supported by the Department of Energy under Grant No. DE-FG02-90ER40542.

In addition to the publishers acknowledged on each individual reprint, I also wish to thank World Scientific for the use of material originally prepared for their volumes commemorating the 50th anniversary of Yang–Mills theory. Chapter 3 on anomalies is largely based on an essay I contributed to *50 Years of Yang–Mills Theory*, edited by G. 't Hooft, and the parts of Chapters 7 and 9 dealing respectively with monopoles and projective group representations are based on an essay I wrote for a projected companion volume on the influence of Yang–Mills theory on mathematics. Also, some material in Chapters 2 and 3 overlaps with the contents of a letter on antecedents of asymptotic freedom that I wrote to *Physics Today*, which appears in the September, 2006 issue.

Contents

REPRINTED PAPERS

1. Early Years, and Condensed Matter Physics

2. High Energy Neutrino Reactions, PCAC Relations, and Sum Rules

4. Quantum Electrodynamics

5. Particle Phenomenology and Neutral Currents

1. Early Years, and Condensed Matter Physics

A brief synopsis of my career appears in an article that I wrote recently for the Abdus Salam International Centre for Theoretical Physics (Adler, 2004, R1), which includes a description of events when I was young that led to my becoming a theoretical physicist. The focus of this article is on the career path that led to my work in high energy physics. However, before I published anything in high energy theory, I spent several summers working in industrial research laboratory jobs in condensed matter physics, and it was this work that led to my first scientific publications.

By the end of my junior year at Harvard, I had taken courses in quantum mechanics and also in condensed matter physics (then called solid state physics). With this background, during the summer of 1960, I got a job working for Joseph Birman, who at that time (before going on to Professorships at New York University and then City College of the City University of New York) headed a section studying electroluminescence at the General Telephone and Electronics (GT&E) Research Laboratory. This industrial research laboratory, formerly the Sylvania Research Laboratory, was conveniently located a few miles from where my family lived in Bayside, Queens. I had a desk in an office looking out over the entrance to the Long Island Sound, from which I could see sections of roadway being hoisted into place on the Throgs Neck Bridge, then under construction.

During my first weeks at GT&E, Joe got me started learning some basic group theory as applied to crystal structures, and then suggested the problem of using these group theory methods to check a formula that Hopfield (1960) had given relating band theory structures in hexagonal and cubic variants of zinc sulfide (ZnS) and related compounds, substances that Joe had been studying (Birman, 1959) with an eye to electroluminescence applications. This turned out to be basically a technical exercise and confirmed Hopfield's results. In the course of this work, which I finally wrote up a year later (Adler, 1962a, R2), I also attempted an *a priori* estimate of a parameter determined by experimental fits to the Hopfield formula. This got me interested in the Ewald sum method for doing crystal lattice sums, on which I wrote a paper (Adler, 1961) giving generalized results for sums over lattices of functions $f(r)Y_{\ell m}(\theta,\phi)$, with $Y_{\ell m}$ a spherical harmonic and $f(r)$ a radial function representable as a transform by $f(r) = r^{\ell}\int_0^{\infty}\exp(-r^2 t)g(t)dt$. These two pieces of work stemming from my summer at GT&E were my first scientific publications. With Joe's encouragement, I also gave a 10 minute contributed paper (Adler and

Birman, 1961) on the ZnS work at the New York meeting of the American Physical Society the following winter, while I was back home on inter-term break from college. Since this was my first conference talk, I typed out a text and went over it so many times that I knew it by heart. After my talk, Joe said words to the effect, "That was fine, but next time you give a talk don't sound like it was memorized", wisdom that I have taken to heart on many subsequent occasions!

When I returned to Harvard for my senior year I was told by some of the faculty that Henry Ehrenreich from the General Electric (GE) Research Laboratory was on leave at Harvard that year, and was giving the graduate course on solid state physics, covering substantially different material from what I had heard the year before. I attended Henry's lectures, which included a calculation of the energy and wave-number dependent dielectric constant in isotropic solids, using the self-consistent field or energy-band approximation, along the lines of the treatment given in Ehrenreich and Cohen (1959). I got to know Henry outside the classroom as well, and he invited me to work at the GE Research Laboratory in Schenectady, NY the following summer, after my graduation from college in June 1961. This was appealing in a number of ways, since my family had moved to Bennington, VT the year before, about an hour's drive away from Schenectady, and so I was able to drive home for a visit on weekends. At GE, Henry suggested that I generalize the treatment of the dielectric constant that he and Cohen had given so as to include various effects of interest in real solids. In the paper that resulted (Adler, 1962b, R3), I calculated the full frequency and wave-number dependent dielectric tensor in the energy-band approximation, including tensor components that couple longitudinal and transverse electromagnetic disturbances, which are absent in the isotropic approximation but are present even in solids with cubic symmetry. The longitudinal to longitudinal component of the general dielectric tensor reduces to the result obtained by Ehrenreich and Cohen when various identities (reflecting charge conservation and gauge invariance, as well as symmetries) are used. I also gave a method, based on an analysis of "Umklapp" processes that couple wave numbers differing by a reciprocal lattice vector, together with use of a multipole expansion, for calculating local field corrections to the dielectric constant, giving a modified Lorenz–Lorentz formula. (Local field corrections were also studied by Cohen's student Nathan Wiser (1963) by a different method.) My paper on the dielectric constant in real solids has been widely cited in the subsequent condensed matter literature, reflecting its relevance for spectroscopic studies of solids, as well as its generalizations to nonlinear dielectric behavior.

Although I had decided to focus on elementary particle theory for my graduate study in Princeton, I retained an interest in solid state physics, and returned to GE for half of the summer of 1962 to work again with Henry Ehrenreich, this time publishing a paper (Adler, 1963) in which I applied the dielectric constant results of the previous summer to the theory of hot electron energy loss in solids. Not long after

this visit, Henry left GE to take a Professorship at Harvard, where our paths crossed again during my postdoctoral years. After finishing my PhD at Princeton in 1964, I spent the summer working at Bell Telephone Laboratories in Murray Hill, under the supervision of Phil Anderson and Dick Werthamer. However, aside from informal notes on the application of raising and lowering operators to the vortex structure in type II superconductors, my principal publication resulting from this final industrial summer job was a writeup of my work on PCAC consistency conditions, which I will discuss in the next chapter.

References for Chapter 1

Adler, S. (1961). A Generalized Ewald Method for Lattice Sums. *Physica* **27**, 1193-1201. [1]

Adler, S. L. (1962a) R2. Theory of the Valence Band Splittings at $k = 0$ in Zinc-Blende and Wurtzite Structures. *Phys. Rev.* **126**, 118-122. [1]

Adler, S. L. (1962b) R3. Quantum Theory of the Dielectric Constant in Real Solids. *Phys. Rev.* **126**, 413-420. [2]

Adler, S. L. (1963). Theory of the Range of Hot Electrons in Real Metals. *Phys. Rev.* **130**, 1654-1666. [2]

Adler, S. L. (2004) R1. From Elements of Radio to Elementary Particle Physics, in *One Hundred Reasons to be a Scientist* (The Abdus Salam International Centre for Theoretical Physics, Trieste), pp. 25-26. [1]

Adler, S. and J. L. Birman (1961). An LCAO Theory of the $\vec{k} = 0,0,0$ Valence Band Splittings in Zinc Blende and Wurtzite Structures. *Bull. Am. Phys. Society* Series II, Vol. 6, No.1, Part 1, p. 22. [1,2]

Birman, J. L. (1959). Simplified LCAO Method for Zincblende, Wurtzite, and Mixed Crystal Structures. *Phys. Rev.* **115**, 1493-1505. [1]

Ehrenreich, H. and M. H. Cohen (1959). Self-Consistent Field Approach to the Many-Electron Problem. *Phys. Rev.* **115**, 786-790. [2]

Hopfield, J. J. (1960). Fine Structure in the Optical Absorption Edge of Anisotropic Crystals. *J. Phys. Chem. Solids* **15**, 97-107. [1]

Wiser, N. (1963). Dielectric Constant with Local Field Effects Included. *Phys. Rev.* **129**, 62-69. [2]

2. High Energy Neutrino Reactions, PCAC Relations, and Sum Rules

Introduction

By the end of my undergraduate years at Harvard (1957-1961), I had gone through most of the graduate course curriculum, as well as a senior year reading course organized by Paul Martin for my classmate Fred Goldhaber and me. This course gave me an introduction to quantum field theory, or more precisely, to quantum electrodynamics, through some of the seminal papers appearing in the reprint volume edited by Schwinger (1958). Although as a result of my summer research jobs I could have gone on relatively easily to a PhD in solid state physics, I wanted to enter particle physics, and moreover wanted exposure to styles of theoretical physics different from those I had seen already at Harvard. Hence I decided on Princeton for my graduate work (with strong encouragement from Harvard faculty member Frank Pipkin, who was an enthusiastic Princeton graduate alumnus), and enrolled there in the fall of 1961.

My first year there was spent preparing for general exams, mostly by reading. I also participated in a seminar organized by the graduate students, which surveyed many aspects of dispersion relations and covered some topics in Feynman diagram calculations as well. The only formal course I took was one given by Sam Treiman, which gave an introductory survey to elementary particle physics. I was impressed by the clarity of his approach, and both because of this and because Murph Goldberger was planning a sabbatical leave the following year, I asked Treiman to take me on as a thesis student.

This turned out to be a fortunate choice. Treiman proposed that I do a thesis in the general area of high energy neutrino reactions, which was just then emerging as an area of phenomenological interest. After doing a survey of the literature in the field, I first did a "preliminary problem" of calculating the final lepton and nucleon polarization effects in the quasielastic neutrino reaction $\nu_\ell + N \to \ell + N$, with all induced form factors retained in the vector and axial-vector vertices (Adler, 1964a). I did this calculation in two ways, first by using the covariant form of the matrix element and Dirac γ matrix algebra, then by using the center of mass form and Pauli matrix algebra, and directly checked the equivalence of the two forms of the answer. This convinced Treiman that I could calculate, and incidentally introduced me to the axial-vector current and coupling g_A which were to be central to my work for many years.

After this calculation was completed, I decided to make the main focus of my thesis a calculation of the simplest inelastic high energy neutrino reaction, that of pion production in the (3,3) or $\Delta(1232)$ resonance region. This problem had the appeal of having as a paradigm the beautiful dispersion relations calculation of pion photoproduction of Chew, Goldberger, Low, and Nambu (1957), which was one of the classics of the dispersion relations program. An extension to electroproduction had already been carried out by Fubini, Nambu, and Wataghin (1958), but they had done no numerical work, and on closer examination their matrix element turned out to be divergent at zero hadronic momentum transfer ν_B when the lepton four-momentum transfer squared denoted by q^2 (or k^2) is nonzero. There were similar problems (surveyed in my thesis) with the other papers then available dealing with pion electroproduction or weak production, so doing a complete and careful calculation, including numerical evaluation of the cross sections, seemed a good choice of thesis topic. It was also a demanding one; although I wrote my thesis and got my degree in 1964, my goal of a complete calculation, including the necessary computer work, was not achieved until 1968.

Much of the delay though, was a result of the fact that weak pion production turned out to be a marvelous theoretical laboratory for studying the implications of conservation hypotheses for the weak vector and axial-vector currents, and this became a parallel part of my research program, as reflected in the title of my thesis "High Energy Neutrino Reactions and Conservation Hypotheses" (Adler, 1964b). From Treiman and from my reading, I had learned about the Feynman–Gell-Mann (1958) proposal of a hadronic conserved vector current (CVC), and I had also learned about the Goldberger–Treiman (1958) relation for the charged pion decay constant, which they had discovered through a pioneering dispersion theoretic calculation of the weak vertex. A simplified derivation of this relation had already been achieved through the suggestion of Nambu (1960), Bernstein, Fubini, Gell-Mann, and Thirring (1960), Gell-Mann and Lévy (1960), and Bernstein, Gell-Mann, and Michel (1960), that the axial-vector current is partially conserved, in the sense that the divergence of the axial-vector current behaves at small squared momentum transfer as a good approximation to the pion field, or equivalently, is pion pole dominated. (Much later on, after contacts with China resumed, I learned that Chou (1960) had given a similar simplified derivation of the Goldberger–Treiman relation, as well as further applications to decay processes.) The partial conservation hypothesis was an appealing one, but as Treiman kept emphasizing, it was supported by "only one number" and therefore had to be regarded with caution. So a second goal of my thesis work ended up being to keep an eye out for other possible tests of the conservation hypotheses for the weak vector and axial-vector currents.

Before going on to discuss how these emerged from my weak pion production calculation, let me first recall what I knew when I started the thesis work. The first chapter of the thesis (written in the spring of 1964) was a theoretical survey; in

the section headed "Partially Conserved Axial Vector Current (PCAC)" I referred only to the papers of Goldberger and Treiman, of Nambu, of Bernstein et al., and of Gell-Mann and Lévy cited in the preceding paragraph. In the final section of the first chapter, entitled "Survey of Computations Relating to Specific Reactions" there is the following reference to the paper of Nambu and Shrauner (1962), which was my reference 37: "An entirely different approach to weak pion production in the low pion-energy region has been pursued by Nambu and Shrauner.[37] These authors assume that the weak interactions are approximately γ_5 invariant ("chirality conservation"). They then obtain formulas for production of low energy pions, in the approximation in which the pion mass is neglected, in analogy with the treatment of low energy bremsstrahling (sic) in electron scattering." At the time I started my calculations, neither Treiman nor I understood the relation between the Nambu–Shrauner work and the issue of partial conservation of the axial-vector current. This was partly because we were suspicious of the assumption of zero pion mass, and partly because the Nambu–Shrauner paper makes no reference to the axial-vector coupling g_A, so it was not clear whether their "chirality" was related to the weak currents I was studying in my thesis. This second point is particularly significant, and I will return to it in considerable detail below. I was not able to determine from my files (by finding either a reference in my notes or a Xerox copy) when I first read the Nambu–Lurié (1962) paper on which the Nambu–Shrauner paper was based, but it was probably a year later, in early 1965.

Forward Lepton Theorem

Roughly the first year and a half of my thesis work on weak pion production was spent mastering the formal apparatus of Lorentz invariant amplitudes (used for writing dispersion relations) and center of mass multipole expansions (used for implementing unitarity) and the transformations between them, the Born approximation structure, cross section calculations, etc. Then in the winter of 1963-1964 or the spring of 1964 (I can only establish dates approximately by the sequence of folders, since I did not date them), I began noticing things that transformed a hard and often dull calculation into a very interesting one (just in the nick of time, since I was due to finish in June of 1964 and had already accepted a postdoctoral position at the Harvard Society of Fellows starting in the fall semester.)

The first thing I noticed was that at zero squared leptonic four momentum transfer, my expression for the weak pion production matrix element reduced to just the hadronic matrix element of the divergence of the axial-vector current, which by the partial conservation hypothesis is proportional to the amplitude for pion-nucleon scattering. I then tried to abstract something more general from this specific observation, and soon had a neat theorem showing that in a general inelastic high energy neutrino reaction, when the lepton emerges forward and the lepton mass is

neglected, the leptonic matrix element is proportional to the four momentum transfer; hence when the leptonic matrix element is contracted with the hadronic part, the vector current contribution vanishes by CVC, and the axial-vector current contribution reduces by partial conservation (for which I coined the parallel acronym PCAC, which has become standard terminology) to the corresponding matrix element for an incident pion. Thus inelastic neutrino reactions with forward leptons can be used as potential tests of CVC and PCAC; this became a chapter of my thesis and was written up as a paper (Adler, 1964c, R4) as soon as my thesis was completed. The paper on CVC and PCAC tests was the first of three papers in which I found connections between high energy neutrino scattering reactions and properties of the weak currents; the other two were my long paper on the g_A sum rule, and a paper on neutrino reaction tests of the local current algebra, both of which are reprinted in this volume and will be discussed shortly.

To determine whether the CVC/PCAC test could be implemented experimentally, I wrote a letter to the neutrino experimentalists at CERN. After a few months I received a charming reply from Gilberto Bernardini, who commented "The delay of this answer, for which I apologize very much, is due to two facts. The first is the known time diagram of the 'modern physicist'. In case you do not know it yet, I plot it here: (Diagram with a vertical time axis and an upwards pointing arrow; 'work' at the bottom, 'travel & meetings' in the middle, and 'dinners & ceremonies' at the top.) Unfortunately, according to my age, I am already very much in the central region and even higher." Bernardini then went on to say that Antonino Zichichi had brought my paper to his attention a couple of weeks before, and then continued with an analysis of technical problems in executing my proposal. There followed a further exchange of letters with Bernardini, with theorist John Bell, and with experimentalists Guy von Dardel and Carlo Franzinetti. Of particular note, von Dardel wrote me a long letter after he read my paper, remarking that the care with which he read it was partly due to a skiing accident that had kept him in bed with a broken leg and nothing better to do, and giving a formula that he had worked out, during his enforced time away from experimental activities, for corrections to my theorem when the lepton emerges at a small angle to the forward direction. This formula turned out to be not quite right (there was an incorrect energy factor), but started me thinking about the issue, which I discussed with John Bell when I attended an Informal Conference on Experimental Neutrino Physics at CERN, January 20-22, 1965. Bell had redone the calculation of the pion exchange contribution to the small angle correction by splitting the amplitude into spin-flip and non-spin-flip parts, getting a result that turned out also to be not quite right (there was a factor of 2 off in one term). When I got back to Harvard I repeated the calculation, according to my notes, by the "Bell method", and also by a covariant method, and got a formula that I never published, but conveyed in letter of Feb. 10, 1965 to Bell (with copies to Bernardini, Block, von Dardel, Faissner, Franzinetti, and Veltman, most of whom

I had talked with when I was at CERN). The corrected small angle formula states that the first factor on the second line of Eq. (16) of R4 should be replaced by

$$\left[1 - \frac{m_\ell^2 k_0}{2k_{20}(k^2 + M_\pi^2)}\right]^2 + \left[\frac{m_\ell k_0 \theta}{2(k^2 + M_\pi^2)}\right]^2 ,$$

with $k^2 = m_\ell^2 k_0/k_{20} + k_{10}k_{20}\theta^2$ the leptonic four-momentum transfer squared and with θ the lepton-neutrino polar angle, assuming that the lepton-neutrino azimuthal angle has been averaged over.

Even before my visit to CERN, Bell (1964) had noted that when one considers my forward lepton formula in the context of nuclei, "the following difficulty presents itself: Because of absorption, pion cross sections depend on the size of large nuclei roughly as $A^{2/3}$. But neutrinos penetrate to all parts of nuclei; for them cross sections should contain at least a part proportional to A. This indicates for large nuclei a critical dependence of $\sigma(W, -q^2)$ on q^2." Bell proceeded to use optical model methods to discuss this "shadowing effect", which has continued to be of interest over the years. It took many years for my forward lepton formula, and Bell's shadowing observation, to be experimentally verified; for a survey of the status of both, and further references, see the recent conference talk by Kopeliovich (2004). An earlier review of Mangano et al. (2001) also discusses the experimental status of shadowing, and a good exposition of the theory is given in the review of Llewellyn Smith (1972). For specific applications of the forward lepton formula to exclusive channels, see Ravndal (1973) and Rein and Sehgal (1981) for $\Delta(1232)$ production, and Faissner et al. (1983) for coherent π^0 production (which was used to determine the coupling strength of the isovector neutral axial-vector current). Also, Sehgal (1988) and Weber and Sehgal (1991) discuss an interesting analog of the forward lepton theorem for purely leptonic neutrino-induced reactions.

Soft Pion Theorems

Returning now to my thesis work in the spring of 1964, the second thing that I noticed, again working from my explicit expression for the weak pion production amplitude, was that when I imposed the PCAC condition at zero values of the hadronic energy variable ν and the hadronic momentum transfer variable ν_B, only the Born approximation pole term coming from the nucleon intermediate state contributed; all of the model dependent parts of the weak amplitudes dropped out. Thus I got what I called a "consistency condition" on the pion-nucleon scattering amplitude $A^{\pi N(+)}$, implied by PCAC, taking the form

$$g_r^2/M = A^{\pi N(+)}(\nu = 0, \nu_B = 0, k^2 = 0)/K^{NN\pi}(k^2 = 0),$$

with g_r the pion-nucleon coupling constant, M the nucleon mass, $-k^2$ the squared mass of the initial pion (the final pion is still on mass shell), and with $K^{NN\pi}(0)$ the

pionic form factor of the nucleon, normalized so that $K^{NN\pi}(-M_\pi^2) = 1$. This seemed absolutely remarkable, and I immediately proceeded to do a dispersion relation evaluation of the pion-nucleon amplitude on the right, using the Roper (1964) phase shift analysis as input, and assuming that the effects of off-shell continuation in $A^{\pi N(+)}$ (as well as in $K^{NN\pi}$) were small. In setting up this calculation, I used several theoretically equivalent ways of writing the subtracted dispersion relation to get an estimate of the errors in the analysis. The Christenson–Cronin–Fitch–Turlay (1964) experiment on CP violation had a substantial block of computer time reserved for analysis, and courtesy of them I was able to use a small amount of their time to run my programs, a few days before I was scheduled to give a talk at Columbia. I recall staying up all night to get the job done, and at one point, in the wee hours of the morning, dropping my deck of cards and then having to spend precious time getting them back in the proper order. But I did get my calculation done by morning (and never again attempted an "all-nighter".) The relation worked very well, and as Treiman later said, "now there is a second number"; PCAC was starting to look interesting. This work became the final chapter of my thesis.

Immediately after finishing my thesis I took a summer job at Bell Laboratories at Murray Hill, nominally working for Phil Anderson. I wanted to learn about superconductivity, and Phil assigned me to work for Dick Werthamer. I did learn about the BCS and Ginzburg–Landau theories, and Abrikosov vortices in type-II superconductors, but I did not succeed in my project with Dick, which was to try to understand the resistance to vortex line motion using thermal Green's functions. A few weeks before the end of the summer, I asked for and got Phil's permission to spend some time writing a paper on the pion-nucleon consistency condition (Adler, 1965a, R5), which I also then extended to pion-pion and pion-lambda scattering. In the pion-pion case, since there are no pole terms, the consistency condition takes the form that the pion-pion scattering amplitude with one zero mass pion, evaluated at the symmetric point $s = t = u = M_\pi^2$, is *zero*. This was the first example of a soft pion zero or, as termed in the literature, "Adler zero", in non-baryonic amplitudes, that I will return to shortly. Knowing that I was planning to go on in particle theory, Phil told me one day that he had an interesting paper to show me, which had just been submitted to the journal *Physics* which he was editing. It was Gell-Mann's (1964) paper on current algebra; Phil let me read it, but not Xerox it. This was to prove decisive for my work on sum rules nine months later. My interactions with Phil however were brief, and never touched on the subject of symmetry breaking in superconductivity and particle physics, on which Phil had written a paper (Anderson, 1963) that I learned of only many years later, that was a forerunner of work on the "Higgs mechanism" for giving masses to vector bosons.

In the fall of 1964 I moved to Harvard as Junior Fellow in the Society of Fellows, sharing a postdoc office next to the office occupied by Henry Ehrenreich in

the Applied Physics division. (Henry had recently left General Electric to accept a Professorship at Harvard.) In principle I was going to do solid state physics as well as particle theory, but that never happened. I spent the fall term working on numerical aspects of my weak pion production calculation, and also reading papers on attempts to calculate the axial-vector renormalization constant g_A, including the papers of Gell-Mann and Lévy (1960) and Bernstein, Gell-Mann and Michel (1960). I had a hunch that the fact that g_A is near one was somehow connected with PCAC, but I did not see a concrete way of exploiting PCAC in a calculation. I also was starting to think about how to make the PCAC consistency condition calculations independent of the cumbersome Lorentz invariant amplitude apparatus that I had used to get them. I soon found that the relevant terms could be isolated directly from the Feynman diagrams without invoking all the formal kinematic apparatus of my thesis, and this approach extended to a general matrix element as well; the strategy was the same one that I had used in the paper on CVC and PCAC tests, of going from a particular observation in the context of my weak pion production calculation to something more general. The result was a formula for soft pion production, in terms of external line insertions on the hadronic amplitude for the same process in the absence of the pion (Adler, 1965b, R6). For baryons of nonzero isospin, the insertion factors are nonzero, while for isospin zero baryons, and mesons such as the pion or kaon, the insertion factor vanishes. This latter result generalizes the soft-pion zero or "Adler zero" to the emission of a soft pion in any reaction involving only incoming and outgoing mesons, but no external baryons. These zeros continue to play a role in the analysis of experimental results on mesonic resonances; for recent discussions, see Bugg (2003, 2004) and Rupp, Kleefeld, and van Beveren (2004).

The soft pion zeros are an indication that according to PCAC, the pion coupling to other hadrons is effectively pseudovector, and not pseudoscalar. When I visited CERN in late January of 1965, while in the midst of work on the Feynman diagram approach to the PCAC consistency conditions, I found that Veltman had been thinking in a similar direction, but had not reached the point of writing down external line insertion rules. Veltman gave me a one page memo to file that he had written, which pointed out that my PCAC consistency conditions are equivalent to pseudovector coupling, which implies the vanishing of invariant amplitudes for soft pion emission after singular terms are split off. Veltman also noted that Feynman had briefly remarked on the relation between the Goldberger–Treiman relation and pseudovector coupling in his conference summary talk at Aix-en-Provence in 1962, and gave me a copy of the relevant page. Feynman did not, however, report agreement with experiments on pion-nucleon scattering, apparently because he did not recognize the necessity of splitting off the singular Born terms before concluding that pion emission amplitudes vanish in the soft pion limit.

In the course of my work on the insertion rules I remembered the paper of Nambu and Shrauner (1962) which I had briefly mentioned in the Introduction to my thesis;

I now looked this up, as well as the paper of Nambu and Lurié (1962) on which it was based, and saw that my final formula, when specialized to the case of an ingoing and outgoing nucleon line, was substantially the same as the pion bremsstrahlung formula of Nambu and Lurié. I noted this in my paper, and consistently referred to the Nambu papers from this point on. In recognition of Nambu's work, I used his notation χ and term "chirality" to refer to the integrated axial-vector charge in my next two papers, which dealt with the g_A sum rule; however, in modern terms this is a misnomer, since chirality is now used to mean the left- or right-handed sums of vector and axial-vector charges. Gell-Mann's notation for the axial-vector charges has become the standard one, and after these two papers I followed the Gell-Mann notation.

The comparison with Nambu's approach also raised the issue of the role of the pion mass: do the PCAC results limit smoothly to the zero pion mass ones, for which the soft pion theorem derivations appear quite different? This point was dealt with in footnote 6 of my paper R6, where I showed that the limits, (1) pion mass approaches zero, and (2) pion four momentum transfer squared approaches zero, can be taken in either order; the same soft pion theorem results, although the contribution which comes from the massless pion pole when the limit (1) is taken first, comes instead from the axial-vector divergence when the limit (2) is taken first. This point is now taken for granted, but in the early years it caused me (and others) considerable confusion. After this paper I almost immediately got involved with sum rules, and so I did not publish the detailed connection between my second PCAC paper and the Nambu–Lurié approach until a few years later, when I included it as "Appendix A" of Chapter 2 of the book on *Current Algebras* which I put together with Roger Dashen (Adler and Dashen, 1968). This appendix is reprinted here as R7. At the end of Appendix A, I again discussed the relationship between the zero pion mass and nonzero pion mass calculations. The analysis of Appendix A also shows how the PCAC approach to soft pion theorems that I had developed fixes the undetermined renormalization constant appearing in the chirality approach of Nambu–Lurié. In the formulas of Appendix A, there are factors of g_A that are missing in the formulas of the papers of Nambu, Lurié, and Shrauner. Correspondingly, in the paper of Nambu and Lurié, in the discussion associated with their Eq. (2.7), they noted that a renormalization constant Z appears, but didn't observe that this can be precisely identified as g_A. Instead, they redefined their chirality as $Z^{-1}\chi$, that is as $(g_A)^{-1}\chi$. They then made a compensating adjustment in the pion decay constant in their Eq. (4.5), where they dropped the g_A factor which appears in the Goldberger–Treiman relation. Nambu and Lurié say there, "$1/\lambda$ is more or less the conventional pion coupling constant $1/\lambda = f = g/2m$. (4.5) It is not proven, however, that this agrees with the coupling constant defined in the dispersion theory. For the time being, we assume it to be the case." In the subsequent paper of Nambu and Shrauner (1962), an identification of f with the standard pion-nucleon coupling was established, but

the issue of where to include factors of g_A was not addressed. My impression from this was that there was some uncertainty in the minds of Nambu and his students about how the chirality is to be normalized, and this impression was reinforced by a conversation I later had with Nambu about their work and my Appendix A derivation of their result.

In the low energy theorem for *one* soft pion which Nambu and Lurié had derived, and which I had obtained from PCAC and the Feynman rules in my second PCAC paper, the g_A factor drops out, and so the normalization of the axial-vector charge or "chirality" is irrelevant. The applications discussed in the papers of Nambu, Lurié, and Shrauner all involved only one soft pion; Nambu and Lurié looked, for example, at $\pi + N \to \pi + N + \pi$, with the final pion soft but with the other pions "hard"; in fact, what they actually did was to calculate single soft pion emission in the reaction $\pi + N \to \Delta(1232)$. Similarly, Nambu and Shrauner (1962) analyzed single soft pion electroproduction and weak production, relating them to the form factors of the vector and axial-vector currents. In this paper they included current commutator terms by analogy with the classic Low (1958) paper on bremsstrahlung; their answer for electroproduction is correct because the g_A factor drops out there anyway, but their answer for weak axial-vector production lacks g_A factors in places, for reasons explained in the next paragraph. A follow-up paper of Shrauner (1963) dealt with single soft pion production in pion-nucleon scattering, with the scattering pions "hard". My "PCAC consistency condition" was likewise a single soft pion theorem which gives a relation between the amplitude for $\pi + N \to N + \pi$, with the final pion soft, and the amplitude $\pi + N \to N$, which is just the pion-nucleon coupling constant, and involves no factors of g_A.

The factors of g_A and the explicit identification of the "chirality" with the charge associated with the axial-vector current become important, however, if one wants to discuss *multiple* soft pion production, and also weak axial-vector pion production, since one then encounters commutators of an axial-vector charge with an axial-vector charge or current, which are evaluated by the Gell-Mann current algebra. If one defines the relevant chirality as $(g_A)^{-1}$ times the axial-vector charge, as is implicit in the Nambu–Lurié paper when one identifies their Z with g_A, then the relevant commutator is $(g_A)^{-2}$ times a vector charge, which at zero momentum transfer just gives $(g_A)^{-2}$. This is in fact the origin of the $(g_A)^{-2}$ term in the g_A sum rule, where the difference between $(g_A)^{-2}$ and 1 is highly significant. The point, then, is that while Nambu and Lurié gave a correct formula for single soft pion production, it in fact cannot be generalized to multiple soft pion production (or soft pion production by the weak axial-vector current) without first dealing carefully with the question of normalization, as I did in my second PCAC paper R6 and in Appendix A of the book on current algebras R7.

Another difference between the work of Nambu and his students, and what I did in my first PCAC consistency condition paper R5, related to the method of

comparison with experiment, and the level of accuracy claimed for soft pion predictions. The Goldberger–Treiman relation is good to about 7% accuracy, and my comparison of the PCAC consistency condition with experiment also indicated that the relation was satisfied to about 10%, thus reinforcing the idea that PCAC could be used as a quantitative tool for studying the strong interactions, with the residual errors arising from the extrapolation of the pion four-momentum squared k^2 from M_π^2 to 0. Given that the pion mass is much smaller than all other hadron masses, an extrapolation error $\sim M_\pi^2/M_{\text{hadron}}^2 \leq 0.1$ is reasonable. The success of the g_A sum rule shortly afterwards gave further support to the idea that PCAC gives quantitatively accurate predictions. Nambu, Lurié, and Shrauner, however, argued only for qualitative agreement between their soft pion results and experiment based on comparisons of rescaled angular distributions, but did not find anything close to $\sim 10\%$ agreement for absolute cross sections. For example, for the relation between the cross sections for pion-nucleon scattering with production of an additional pion, and pion-nucleon scattering, Nambu and Lurié (1962) showed agreement with their predictions to within roughly a factor of three (giving a predicted cross section of 0.2 mb versus experimental values in the range 0.6 to 0.7 mb). Similarly, for the same reaction Shrauner (1963) found that "the magnitudes of the cross sections seem to be significantly underestimated by a factor of about 7". The source of these discrepancies is not clear. They may be due, in part, to the fact that, instead of testing the soft pion predictions at the kinematic point of zero pion four momentum (such as the point $\nu = \nu_B = 0$ used in my PCAC consistency condition work), Nambu, Lurié, and Shrauner did the comparisons in energy intervals above scattering threshold. (However, Shrauner argues, on the basis of branching ratios,that the discrepancy is probably not attributable to an overlap of the $\Delta(1232)$ resonance with the comparison region.) I think that a combination of lack of clarity about how their chiral current was related to the physical axial-vector current, as reflected in the normalization problems noted above, together with the lack of striking quantitative comparisons with experiment, were responsible for the work of Nambu and his students being largely unnoticed by the community. It was only after the quantitative successes of PCAC in my consistency condition paper and in the g_A sum rule that followed shortly afterwards, and my demonstration of the equivalence between the PCAC insertion rules and the chirality conservation approach, that the significance of the work of the Nambu group became clear.

Finally, as an historical footnote to this discussion of soft pion theorems, Touschek (1957) appears to have been the first to introduce continuous γ_5 symmetry transformations, as applied to the neutrino field, and to observe that invariance under these transformations requires that the neutrino mass be zero. Nishijima (1959) (in work submitted for publication in late 1958) considered continuous γ_5 symmetry transformations in theories of massive fermions; to preserve γ_5 invariance he gauged the transformations with a massless pseudoscalar boson, transforming as $B \rightarrow B + \lambda$

under a γ_5 transformation with parameter λ. The action written in Nishijima's paper is just the effective action one would now write for a singlet Nambu–Goldstone boson (such as an axion) coupled to a massive fermion. Nambu (1959), in remarks at the Kiev Conference, noted the analogy between γ_5 symmetry in particle physics and gauge invariance in superconductivity, and related this to his suggestion that a nucleon-antinucleon pair in a pseudoscalar state could be the pion. This idea was further developed in the well-known paper Nambu and Jona-Lasinio (1961), that laid the basis for the modern theory of Nambu–Goldstone bosons associated with spontaneous symmetry breaking, and for the fact that most of the mass of the nucleon comes from chiral symmetry breaking. In the meantime, Gürsey (1960) had introduced isovector γ_5 transformations, as an extension of the similar isoscalar transformations used by Nishijima, and constructed a precursor to nonlinear pion effective Lagrangians. These papers all contained important seeds of our present-day understanding of chiral symmetries.

Sum Rules

I have now gotten ahead of the chronological story; a lot of things happened very fast in 1965. In the fall of 1964 I started thinking about the question of the renormalization of the nucleon axial-vector coupling g_A, and accumulated a file of papers on the subject. However, my attempts at a calculation were based on the commutator of the nucleon field with the weak axial-vector charge, giving results identical to those already obtained by Bernstein, Gell-Mann, and Michel (1960), which expressed g_A in terms of unmeasurable off-shell form factors, but achieving no further progress. In early 1965 I saw a preprint of Fubini and Furlan (published as Fubini and Furlan, 1965) which applied the commutator of vector current charges, together with the ingenious idea of going to an infinite momentum frame, to calculate the radiatively induced renormalization of the vector current. (Harvard did not have a preprint library in those days, but Schwinger's secretary Shirley would let me into his office from time to time to look through the unread preprints that were stacked on his desk. This presented no difficulty since Schwinger was a night-owl who mainly worked at home, and used his office only a few hours a week, when he came in to lecture and to see students. That is how I became aware of the Fubini–Furlan paper. As a result of this experience, one of the first things I did when I arrived at the Institute for Advanced Study eighteen months later was to start a preprint library for the particle physicists.) I immediately thought about applying this to the axial-vector current, using the Gell-Mann current algebra that I'd seen the previous summer at Bell Labs. However, because of other things I was working on I didn't get around to it until a few months later, when in a chance encounter Arthur Jaffe told me that he had heard a talk by Roger Dashen about work he and Gell-Mann had been doing on sum rules. I decided I had better stop delaying (although it turned

out that Dashen and Gell-Mann were working on fixed momentum transfer sum rules), dropped my weak pion production computer work, and spent spring break working out the consequences of combining the Gell-Mann current algebra, PCAC, and the Fubini–Furlan method. It turned out to be surprisingly easy, with the infinite momentum frame solving a problem I had encountered in earlier attempts to calculate g_A, which is that the axial-vector charge matrix element is proportional to the nucleon velocity, and vanishes for nucleons at rest. I soon had a formula relating the difference between 1 and $(g_A)^{-2}$ to a convergent integral over a difference of pion-nucleon cross sections,

$$1 - \frac{1}{g_A^2} = \frac{4M_N^2}{g_r^2 K^{NN\pi}(0)^2} \frac{1}{\pi} \int_{M_N+M_\pi}^{\infty} \frac{WdW}{W^2 - M_N^2} [\sigma_0^+(W) - \sigma_0^-(W)] \quad ,$$

with M_π and M_N the pion and nucleon masses, $\sigma_0^\pm(W)$ the total cross section for scattering of a zero-mass $\pi^\pm$ on a proton at center-of-mass energy W, and again with $K^{NN\pi}(0)$ the pionic form factor of the nucleon, normalized so that $K^{NN\pi}(-M_\pi^2) = 1$. I first tried to saturate the integral in the narrow $\Delta(1232)$ approximation, and the result was a disappointing $g_A = 3$. I then pulled out the computer deck I had used for the consistency condition numerical work the previous year, did the integral carefully, and got $g_A = 1.24$. I also observed that the relation for g_A could be equivalently recast as a two-soft pion low energy theorem,

$$1 - \frac{1}{g_A^2} = \frac{-2M_N^2}{g_r^2 K^{NN\pi}(0)^2} G(0,0,0,0) \quad ,$$

$$G(\nu, \nu_B, M_\pi^i, M_\pi^f) = \nu^{-1} A^{\pi N(-)}(\nu, \nu_B, M_\pi^i, M_\pi^f) + B^{\pi N(-)}(\nu, \nu_B, M_\pi^i, M_\pi^f) \quad .$$

Here $A^{\pi N(-)}$ and $B^{\pi N(-)}$ are the isospin-odd pion-nucleon scattering amplitudes, ν and ν_B are again the energy and momentum transfer variables, and $M_\pi^{i,f}$ are the initial and final pion masses, which are now both off shell. A few days after I submitted a letter to *Physical Review Letters*, Sidney Coleman returned from a trip to SLAC and when I described my results to him, he told me that he had just heard about a similar calculation being done there by Bill Weisberger, whose points of departure were the same as mine: the Gell-Mann current algebra, the Fubini–Furlan paper, and my paper on PCAC consistency conditions. I talked to Weisberger by phone, and then called PRL and asked them to delay publication of my letter until they received the manuscript Weisberger was preparing. My paper (Adler, 1965c, R8) and Weisberger's (Weisberger, 1965) appeared as back-to-back letters in the June 21 issue. They give substantially identical derivations; Weisberger's numerical result of 1.16 differed from mine of 1.24 because I had included a correction for the off-pion-mass-shell extrapolation of the threshold phase space factor associated with the $\Delta(1232)$ resonance, which I knew from my work on weak pion production could be reliably estimated. At the time, this correction made agreement with experiment worse (the experimental value for g_A was then 1.18), but the best value now has

settled down to $g_A = 1.257 \pm .003$, in gratifyingly good agreement with the value I got when I included the kinematic extrapolation correction. Weisberger and I both submitted longer papers to *Physical Review* describing our work (Adler, 1965d, R9); Weisberger (1966). These emphasized the low energy theorem approach to the relation for g_A, giving historically the first two-soft pion low energy theorem. In my paper I also gave an analog for pion-pion scattering, and then in the final section (Adler, 1965, R9, Section V), I returned to the observation that I had made a year earlier about forward lepton scattering, and showed that the g_A sum rule could be converted to an *exact* relation, involving no off-shell PCAC extrapolation, for forward inelastic high energy neutrino reactions. This relation, which provided a test of the Gell-Mann current algebra of axial-vector charge commutators, was another indication of a deep connection between the structure of currents on the one hand, and inelastic lepton scattering on the other.

The g_A sum rule provided yet a third result supporting the use of PCAC as a method for calculating soft pion processes. Simultaneously, it was a stunning success for Gell-Mann's brilliant idea of abstracting the current algebra from the naive quark model, with the hope that it would prove to be a feature that would also be valid in the then unknown theory of the strong interactions. At this point the whole community took notice, and a string of current algebra/PCAC applications appeared in rapid succession. To mention just a few, Weinberg (1966a) and Tomozawa (1966) reexpressed the soft pion theorems for pion-nucleon scattering, coming from my consistency condition papers and the g_A sum rule papers, in the form of formulas for the pion-nucleon scattering lengths, and Weinberg in the same paper also used my result of a PCAC zero in pion-pion scattering, plus a symmetry argument, as inputs for a derivation of pion-pion scattering lengths. Weinberg (1966b) also generalized the two-soft pion low energy form of the g_A sum rule to a general formula for multiple soft pion production. Finally, in another striking application of soft pion theorems, Callan and Treiman (1966) gave a series of important results for K meson decays, in which the role of rapidly varying pole terms was clarified in Weinberg (1966c).

In connection with the g_A sum rule, I have an interesting Feynman anecdote to relate. I spent the spring term of 1966 as a member of Murray Gell-Mann's postdoctoral group at Cal Tech. A few weeks after I arrived, Feynman asked me to stop by his office to look at some pages in his notebook, in which he had almost derived the g_A sum rule, before Weisberger and I did it. The whole expression was there (including the kinematic correction that I had included for the off-mass-shell extrapolation), except that, where the Gell-Mann algebra had dictated a 1 coming from the commutator of two axial-vector charges giving an unrenormalized vector charge, Feynman had put 0! So numerically the relation did not work, and Feynman had given up on it and gone on to other things. He evidently had not paid attention to Gell-Mann's current algebra, or at least not realized, from his heuristic way of doing things, that it was essential for this calculation.

Returning again to events in 1965, as soon as the long paper on g_A was completed, I departed to be a summer visitor at CERN. There I met Murray Gell-Mann for the first time, and had long conversations with him. Murray was particularly interested in the Section V relation between the current algebra of vector and axial-vector charges and forward high energy neutrino reactions, and urged me to try to extend it to a test of the *local* current algebra which he had given in his *Physics* paper (Gell-Mann, 1964). I spent the summer working on this, and found that I could do it; as I recall, the crucial bits came together when I spent a day working at a kitchen table during a week off for holiday at Lake Garda. The results were written up in the late summer of 1965 at CERN and/or Harvard, and appeared in Adler (1966), R10. This article gave the first detailed working out of the structure of deep inelastic high energy neutrino scattering (the electroproduction case was given independently in the review of de Forest and Walecka (1966)), with both the electroproduction and neutrino cases specific examples of general local lepton coupling theorems given by Lee and Yang and by Pais, as referenced in my article R10. However, the α, β, γ notation that I used for the structure functions did not become the standard one; the now standard $W_{1,2,3}$ structure functions, which follow the notation of de Forest and Walecka and were further popularized by Bjorken, are linearly related to the ones I used. [Specifically, I separated the cross section into strangeness-conserving and strangeness-changing pieces, whereas the current convention is to define the structure functions as the sum of both. At zero Cabibbo angle, the relation between my α, β, γ and the conventional $W_{1,2,3}$ is $\alpha = W_1$, $\beta = W_2$, $2M_N\gamma = W_3$, with M_N the nucleon mass. For general Cabibbo angle θ_C, one has $\cos^2\theta_C\beta^{(+,-)}_{\Delta S=0} + \sin^2\theta_C\beta^{(+,-)}_{|\Delta S|=1} = W_2^{\nu,\bar{\nu}}$, with similar relations for the other two structure functions.] The article actually gave three sum rules; two for the α and γ structure functions which subsequent analysis by Dashen showed to be divergent and hence useless, and one for the β deep inelastic amplitude which is a convergent and useful relation. The beta sum rule divides into axial-vector and vector parts, which are separately given as Eqs. (53a) and (53b) respectively of Adler (1966), R10, and which when added to give the total $\Delta S = 0$ cross section yield

$$2 = g_A(q^2)^2 + F_1^V(q^2)^2 + q^2F_2^V(q^2)^2 + \int_{M_N+M_\pi}^{\infty} \frac{W}{M_N} dW[\beta^{(-)}(q^2, W) - \beta^{(+)}(q^2, W)] \quad .$$

This sum rule (and the ones for the separate vector and axial-vector contributions) has the notable feature that the left-hand side is independent of q^2, even though the Born term contributions and the continuum integrand on the right are q^2-dependent. At zero squared momentum transfer q^2, the axial-vector part of the β sum rule reduces to the relation I gave in my long paper on g_A, which had prompted Gell-Mann's question about a generalization; the first derivative of the vector part with respect to q^2 at $q^2 = 0$ gives the sum rule also derived by Cabibbo and Radicati (1966) using moments of currents. Because the neutrino and antineutrino differential

cross sections $d^2\sigma/d(q^2)dW$ are dominated by the β structure function in the limit of large neutrino energy, by integrating over W one gets the limiting cross section relation (at zero Cabibbo angle)

$$\lim_{E_\nu\to\infty}\left[\frac{d\sigma(\overline{\nu}+p)}{d(q^2)}-\frac{d\sigma(\nu+p)}{d(q^2)}\right]=\frac{G^2}{\pi}\quad ,$$

with G the Fermi constant. Similar relations at non-zero Cabibbo angle are given in Eq. (27) of R10, and it is easy to obtain analogous relations for the vector and axial-vector contributions to the cross sections taken separately.

In late October of 1965 I spoke on "High Energy Semileptonic Reactions" at the International Conference on Weak Interactions held at Argonne National Laboratory (Adler, 1965e), in which I gave the first public presentation of the local current algebra sum rules for the β deep inelastic neutrino structure functions, and the limiting relations for the differential cross sections that they imply. In the published discussion following this talk, in answer to a question by Fubini, I noted that the β sum rule had been rederived by Callan (unpublished) using the infinite momentum frame limiting method, but that the α and γ sum rules could not be derived this way, reinforcing suspicions that "the integral for β is convergent, while the other two relations (for α and γ) really need subtractions." Bjorken was in the audience and was intrigued by the β sum rule results, and soon afterwards converted them into a differential cross section inequality (Bjorken, 1966, 1967) for deep inelastic electron-nucleon scattering, for which there was the prospect of experimental tests relatively soon. To see why the neutrino cross section relation given above implies an inequality for electron scattering, one notes that since the $\nu+p$ differential cross section is positive, the right-hand side G^2/π gives a lower bound for the $\overline{\nu}+p$ differential cross section, with a similar lower bound holding for the vector current contribution alone. But noting that according to CVC, the vector weak current is in the same isospin multiplet as the isovector part of the electromagnetic current, and using the Wigner–Eckart theorem, one gets a corresponding lower bound for the inelastic differential cross section induced by an isovector virtual photon scattering on a nucleon. One then notes that in the scattering of a virtual photon on a target containing equal numbers of neutrons and protons, the isovector and isoscalar currents add incoherently, and so the isovector current contribution alone gives a lower bound. Combining the two bounds, and including an extra $1/(k^2)^2$ for the virtual photon propagator, replacing G by the fine structure constant α, and keeping track of numerical factors, one gets Bjorken's electron scattering result

$$\lim_{E_e\to\infty}\frac{d[\sigma(e+p)+\sigma(e+n)]}{d(k^2)}>\frac{2\pi\alpha^2}{(k^2)^2}\quad ,$$

which was testable in the experiments soon to begin at SLAC. Verification of my neutrino sum rule, on the other hand, took two decades and more; see Allasia et al. (1985) for the first reported test, and Conrad, Shaevitz, and Bolton (1998) for

more recent high precision results. For a recent study of my neutrino sum rule, in comparison with the Gottfried (1967) sum rule for electron-proton scattering, within the framework of the large N_c expansion of QCD with N_c colors, see Broadhurst, Kataev, and Maxwell (2004) and Kataev (2004).

Although not directly tested until many years after it was derived in 1965, my neutrino sum rule had important conceptual implications that figured prominently in developments over the next few years. To begin with, it gave the first indications that deep inelastic lepton scattering would give information about the local properties of currents, a fact that at first seemed astonishing, but which turned out to have important extensions. Secondly, as noted by Chew in remarks at the 1967 Solvay Conference (Solvay, 1968), the closure property tested in the sum rules, if verified experimentally, would suggest the presence of elementary constituents inside hadrons. In a Letter (Chew, 1967) published shortly after this conference, Chew argued that my sum rule, if verified, would rule out the then popular "bootstrap" models of hadrons, in which all strongly interacting particles were asserted to be equivalent ("nuclear democracy"). In his words, "such sum rules may allow confrontation between an underlying local spacetime structure for strong interactions and a true bootstrap. The pure bootstrap idea, we suggest, may be incompatible with closure." In a similar vein, Bjorken, in his 1967 Varenna lectures (Bjorken, 1968), argued that the neutrino sum rule was strongly suggestive of the presence of hadronic constituents, and this was also noted in the review of Llewellyn Smith (1972).

These conceptual developments still left undetermined the mechanism by which the neutrino sum rule, and Bjorken's electron scattering inequality, could be saturated at large q^2. During my visit to Cal Tech in 1966, I renewed my graduate school acquaintance with Fred Gilman and worked with him on two projects. One was an analysis of the saturation of the neutrino sum rule for small q^2 (Adler and Gilman, 1967, R11), in which we concluded that SLAC (soon to start operating) would have enough energy to confront the saturation of the nonzero q^2 sum rules in a meaningful way. In this paper, we noted that the β sum rule posed what at the time was a puzzle: the left-hand side of the sum rule is a constant, while the Born terms on the right are squares of nucleon form factors, which vanish rapidly as the momentum transfer q^2 becomes large. The low lying nucleon resonance contributions on the right were expected to behave like the $\Delta(1232)$ contribution, which is form factor dominated and also falls off rapidly with q^2. Hence it was clear that something new and interesting must happen in the deep inelastic region if the sum rule were to be satisfied for large q^2: "to maintain a constant sum at large q^2, the high W states, which require a large E to be excited, must make a much more important contribution to the sum rules than they do at $q^2 = 0$". We were cautious, however (too cautious, as it turned out!), and did not attempt to model the structure of the deep inelastic component needed to saturate the sum rule at large

q^2. Bjorken became interested in the issue of how the sum rule could be saturated, and formulated several preliminary models that (in retrospect) already had hints of the dominance of a regime where the energy transfer ν grows proportionately to the value of q^2. I summarized these pre-scaling proposals of Bjorken in the discussion period of the 1967 Solvay Conference (Solvay, 1968), which Bjorken did not attend, in response to questions from Chew and others as to how the neutrino sum rule could be saturated. The precise saturation mechanism was clarified (to a very good first approximation) some months after the Solvay conference with the proposal by Bjorken (Bjorken, 1969) of scaling, and soon afterwards, with the experimental work at SLAC on deep inelastic electron scattering, that confirmed Bjorken's intuition. For a very clear exposition of the relation between scaling and the neutrino sum rule, see Sec. 3.6B of Llewellyn Smith (1972), who notes that when the sum rule is rewritten in terms of Bjorken's scaling variable ω, "The simplest way to ensure the Q^2 [my q^2] independence of the left-hand side as $Q^2 \to \infty$ is to assume that the limit in eq.(3.71) [in my notation, $\lim_{Q^2\to\infty,\omega\,\mathrm{fixed}} \beta^{(\pm)}(\omega, Q^2/M_N^2)$] exists".

More Low Energy Theorems; Weak Pion Production Redux

In the fall of 1965 I received an invitation from Oppenheimer, which I accepted, to come to the Institute for Advanced Study as a long term member with a five year appointment, starting in the fall of 1966. Roger Dashen, whom I had met briefly when he visited Harvard earlier in 1965, received a similar invitation. The intent behind our appointments was that we would reinvigorate high energy theory at the Institute, which had fallen into a decline with the departures of Lee, Yang, and Pais to professorships elsewhere, and with a turn of Dyson's research interests towards astrophysics.

Before going to Princeton, as mentioned above, I spent the spring term of 1966 as a postdoc in Murray Gell-Mann's group at Cal Tech. By this time the successes of PCAC and current algebra had attracted a lot of attention and stimulated an outpouring of papers, the more important ones of which appear in the volume which Dashen and I put together a year later. My own work in the spring of 1966 was focused on two issues. The first involved using PCAC to get small momentum expansions of matrix elements of the axial-vector current, in analogy with the paper of Low (1958) on soft photon bremsstrahlung. With Joe Dothan, I wrote a long paper (Adler and Dothan, 1966, R12) applying these ideas to the weak pion production amplitude and to radiative muon capture. The weak pion results figured in my later comprehensive paper on the subject (see below), while the radiative muon capture work was incorporated into later chiral perturbation theory treatments of radiative muon capture; for a review of the current theoretical and experimental status of muon capture, including a discussion of discrepancies between theory and experiment in the radiative capture case, see Gorringe and Fearing (2004). The other

direction of work involved two phenomenological studies done with Fred Gilman. One of these dealt with saturation of the neutrino sum rule, as described in the preceding section. The other dealt with a detailed phenomenological study of the PCAC predictions for pion photo- and electro-production (Adler and Gilman, 1966, R13), including a saturation analysis for the Fubini–Furlan–Rossetti (1965) sum rule; for a recent update on this, see Pasquini, Drechsel, and Tiator (2005).

My first year at the Institute was largely devoted to writing the book on *Current Algebras* with Roger Dashen (Adler and Dashen, 1968). The book consisted of selected reprints grouped by categories with commentaries that we supplied, plus some general introductory material. I was responsible for writing the introductory sections and the commentaries for Chapters 1-3, which included Appendix A, reprinted here as R7. Roger was responsible for the commentaries for Chapters 4-7, which included an original and very detailed analysis of precisely which sum rules could be derived by the infinite momentum frame method, or in different language, when a naive assumption of unsubtracted dispersion relations would (and would not) give correct results. This analysis confirmed earlier suspicions that my β neutrino sum rule was correct, but that the α and γ sum rules should have subtractions, and so were not useful. The book on *Current Algebras* was completed, and sent off to the publisher, in the fall of 1967.

During this period I also worked with Bill Weisberger, who was then at Princeton, on sorting out the tricky pion pole structure in two pion photo- and electroproduction, which had to be handled carefully to get a fully gauge-invariant expression (Adler and Weisberger, 1968, R14). Our interest in this process, as noted in the title of the paper, was motivated by the fact that it gives an alternative, indirect method of measuring the nucleon axial-vector form factor $g_A(k^2)$. An experiment to measure $g_A(k^2)$ by this method was carried out by Joos et al. (1976) giving a value $m_A = 1.18 \pm 0.07$ GeV for the mass in the dipole formula $g_A(k^2) = g_A(0)(1 + k^2/m_A^2)^{-2}$. This value is in good agreement with the value $m_A = 1.07 \pm 0.06$ GeV given in the quasielastic scattering $\nu_\mu + n \rightarrow \mu^- + p$ experiment of Baker et al. (1981), and also in reasonable agreement with values of m_A obtained from single pion electroproduction at threshold using the low energy theorem of Nambu and Shrauner (1962) (for which experimental references are given in both the Joos et al. and Baker et al. articles). At the 1968 Nobel Symposium on Elementary Particle Theory, I gave a brief talk (Adler, 1968a) reviewing various methods that had been proposed to measure the nucleon axial-vector form factor: quasielastic neutrino scattering, neutrino production of the $\Delta(1232)$, electroproduction of a single soft pion (the Nambu–Shrauner proposal), and electroproduction of the $\Delta(1232)$ plus an additional soft pion (the proposal of my paper R14 with Weisberger). Over the years since then, all of these methods have been carried out.

I also returned, after completion of the book on *Current Algebras*, to the repeatedly delayed project of completing the numerical work associated with my thesis

calculation of weak pion production, and this kept me busy until the spring of 1968, when I finished a comprehensive article on photo-, electro-, and weak single-pion production in the $\Delta(1232)$, or as it was then termed, the (3,3) resonance region (Adler, 1968b, R15). This paper is so long (123 pages) that it is not feasible to reprint it all here, so I have included only the introduction (Sec. 1) and part of the discussion of implications of PCAC (Secs. 5A and 5B). The basic approximation used in this paper consisted of using the Born approximation for all nonresonant multipoles, augmented by terms coming from the PCAC low energy theorems, together with a unitarized Born approximation for the dominant resonant (3,3) multipoles, giving predictions for weak pion production in the (3,3) region in terms of the vector and axial-vector form factors of the nucleon. By 1968 there were experimental results on pion electroproduction which were in satisfactory agreement with my theory, except for values of the momentum transfer k^2 significantly larger than roughly $0.6(\mathrm{GeV}/c)^2$, where in retrospect one can see effects from the scaling regime showing up. For neutrino pion production, preliminary comparison of my results with CERN data showed an axial-vector form factor $g_A(k^2)$ that falls off more slowly with k^2 than the vector form factors, with a dipole mass of $m_A \sim 1.2\mathrm{GeV}$. A subsequent comparison of my model with high-statistics neutrino data from Brookhaven by Kitagaki et al. (1986) gave good fits with a dipole mass of $m_A = 1.28 \pm 0.11$ GeV, somewhat high compared to values obtained by other methods described above. Reasonable fits of my model to the Δ cross section and density matrix elements measured in the hydrogen bubble chamber at Argonne were also reported in papers of Schreiner and von Hippel (1973a,b), and a comparison with other models and data was given by Rein and Sehgal (1981). (For a recent alternative approach to $\Delta(1232)$ weak production, and extensive references to earlier theoretical and experimental studies of this reaction, see Paschos et al. (2004).) After 1968 I did not work again on weak pion production until 1974-75, when the subject became important because it was an avenue for exploring weak neutral currents, as discussed in Chapter 5 below.

To conclude this section on low energy theorems, let me address the question of the extent to which the modern viewpoint, of pions as Nambu–Goldstone bosons, entered into my work. The earliest reference that I could find in my research notes to the "Goldstone theorem" (and specifically to the derivations given in the paper of Goldstone, Salam, and Weinberg, 1962) dates from the spring of 1967, in other words, after nearly all the work on soft pion theorems was completed. (This reference was in the context of calculations on the axial-vector vertex in QED that were the starting point of my work on the axial anomaly, to be discussed in the next chapter.) I fully appreciated the role of pions as Nambu–Goldstone bosons only after hearing seminars that referred to Nambu–Goldstone versus Wigner–Weyl representations of γ_5 symmetry, which were connected (as best I recall) with the work of Gell-Mann, Oakes, and Renner (1968) and Dashen (1969) on chiral $SU(3) \times SU(3)$ as a strong interaction symmetry. This may at first seem surprising, but now that the tapestry

of the standard model is completed, we see clearly the interrelations of its many threads; at the time when these threads were being laid down, those working from one direction were often unaware or only dimly aware of progress from another.

Perhaps this is also a good point to say that the elucidation of the chiral structure of the strong interactions was only *one* of the results flowing from the successes of current algebra methods and PCAC; something that was perhaps even more significant at the time was the demonstration that quantum field theory methods were really valid, after all, in dealing with hadronic interactions. When I entered graduate school, the prevailing view was that the strong interactions would be understood through some kind of dispersion theoretic "reciprocal bootstrap", and nearly every particle physics talk I heard began with a Mandelstam diagram on the blackboard. By 1967, this view had changed; it was clear that field theory could produce results which could not be obtained from the dispersion relations program, and this strongly influenced subsequent developments.

References for Chapter 2

Adler, S. L. (1964a). Polarization Effects in High-Energy Weak Interactions. *Nuovo Cimento* **30**, 1020-1039. [4]

Adler, S. L. (1964b). High Energy Neutrino Reactions and Conservation Hypotheses, Princeton University dissertation, on deposit with University Microfilms. [5]

Adler, S. L. (1964c) R4. Tests of the Conserved Vector Current and Partially Conserved Axial-Vector Current Hypotheses in High-Energy Neutrino Reactions. *Phys. Rev.* **135**, B963-B966. [7,8]

Adler, S. L. (1965a) R5. Consistency Conditions on the Strong Interactions Implied by a Partially Conserved Axial-Vector Current. *Phys. Rev.* **137**, B1022-B1033. [9,12,13]

Adler, S. L. (1965b) R6. Consistency Conditions on the Strong Interactions Implied by a Partially Conserved Axial-Vector Current. II. *Phys. Rev.* **139**, B1638-B1643. [10-12]

Adler, S. L. (1965c) R8. Calculation of the Axial-Vector Coupling Constant Renormalization in β Decay. *Phys. Rev. Lett.* **14**, 1051-1055. [15,16]

Adler, S. L. (1965d) R9. Sum Rules for the Axial-Vector Coupling-Constant Renormalization in β Decay. *Phys. Rev.* **140**, B736-B747. [16,17]

Adler, S. L. (1965e). High Energy Semileptonic Reactions, in *Proceedings of the International Conference on Weak Interactions*, held at Argonne National Laboratory, October 25-27, 1965, ANL-7130, pp. 285-303 (talk) and pp. 304-309 (discussion). [18]

Adler, S. L. (1966) R10. Sum Rules Giving Tests of Local Current Commutation Relations in High-Energy Neutrino Reactions. *Phys. Rev.* **143**, 1144-1155. [17-19,21]

Adler, S. L. (1968a). Measurement of the Nucleon Axial-Vector Form Factor, in *Elementary Particle Theory, Relativistic Groups and Analyticity*, Proceedings of the Eighth Nobel Symposium, held May 19-25, 1968, N. Svartholm, ed. (Almqvist & Wiksell, Stockholm, and John Wiley & Sons, New York), pp. 263-268. [21]

Adler, S. L. (1968b) R15. Photo-, Electro-, and Weak Single-Pion Production in the (3,3) Resonance Region. *Ann. Phys.* **50**, 189-311. Pages 189-192 and 255-266 are reprinted here. [21,22,30,68-70]

Adler, S. L. and R. F. Dashen (1968), Appendix A R7. *Current Algebras and Applications to Particle Physics* (W. A. Benjamin, New York), pp. 139-146. [11,12,21]

Adler, S. L. and Y. Dothan (1966) R12. Low-Energy Theorem for the Weak Axial-Vector Vertex. *Phys. Rev.* **151**, 1267-1277. [20]

Adler, S. L. and F.J. Gilman (1966) R13. Partially Conserved Axial-Vector Current Restrictions on Pion Photoproduction and Electroproduction Amplitudes. *Phys. Rev.* **152**, 1460-1467. [21]

Adler, S. L. and F. J. Gilman (1967) R11. Neutrino or Electron Energy Needed for Testing Current Commutation Relations. *Phys. Rev.* **156**, 1598-1602. [19-21]

Adler, S. L. and W. I. Weisberger (1968) R14. Possible Measurement of the Nucleon Axial-Vector Form Factor in Two-Pion Electroproduction Experiments. *Phys. Rev.* **169**, 1392-1397. [21]

Allasia, D. et al. (1985). Q^2 Dependence of the Proton and Neutron Structure Functions from Neutrino and Antineutrino Scattering on Deuterium. *Z. Phys. C. - Particles and Fields* **28**, 321-333. [18]

Anderson, P. W. (1963). Plasmons, Gauge Invariance, and Mass. *Phys. Rev.* **130**, 439-442. [9]

Baker, N. J. et al. (1981). Quasielastic Neutrino Scattering: A Measurement of the Weak Nucleon Axial-Vector Form Factor. *Phys. Rev. D* **23**, 2499-2505. [21]

Bell, J. S. (1964). Nuclear Optical Model for Virtual Pions. *Phys. Rev. Lett.* **13**, 57-59. [8]

Bernstein, J., S. Fubini, M. Gell-Mann, and W. Thirring (1960). On the Decay Rate of the Charged Pion. *Nuovo Cimento* **17**, 757-766. [5,6]

Bernstein, J., M. Gell-Mann, and L. Michel (1960). On the Renormalization of the Axial Vector Coupling Constant in β-Decay. *Nuovo Cimento* **16**, 560-568. [5,6,10,14]

Bjorken, J. D. (1966). Inequality for Electron and Muon Scattering from Nucleons. *Phys. Rev. Lett.* **16**, 408. [18]

Bjorken, J. D. (1967). Inequality for Backward Electron- and Muon-Nucleon Scattering at High Momentum Transfer. Phys. Rev. **163**, 1767-1769. [18]

Bjorken, J. D. (1968). Current Algebra at Small Distances, in *Proceedings of the International School of Physics "Enrico Fermi" Course XLI*, J. Steinberger, ed., Academic Press, New York, pp. 55-81. See p. 56 of this proceedings. [19]

Bjorken, J. D. (1969). Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.* **179**, 1547-1553. [20]

Broadhurst, D. J., A. L. Kataev, and C. J. Maxwell (2004). Comparison of the Gottfried and Adler Sum Rules Within the Large-N_c Expansion. *Phys. Lett. B* **590**, 76-85. [19]

Bugg, D. V. (2003). Comments on the σ and κ. *Phys. Lett. B* **572**, 1-7. [10]

Bugg, D. V. (2004). Four Sorts of Meson. *Physics Reports* **397**, 257-358; see Sec. 11. [10]

Cabibbo, N. and L. A. Radicati (1966). Sum Rule for the Isovector Magnetic Moment of the Nucleon. *Phys. Lett.* **19**, 697-699. [17]

Callan, C. G. and S. B. Treiman (1966). Equal Time Commutators and K-Meson Decays. *Phys. Rev. Lett.* **16**, 153-157. [16]

Chew, G. F. (1967). Closure, Locality, and the Bootstrap. *Phys. Rev. Lett.* **19**, 1492-1495. Chew also refers to a more general local current algebra sum rule submitted for publication by Fubini after my Argonne Conference presentation, the integrand of which is not expressible in terms of measurable stucture functions: see S. Fubini, Equal-Time Commutators and Dispersion Relations, *Nuovo Cimento A* **43**, 475-482 (1966). [Chew however, through an apparent misunderstanding, gives as the reference an earlier paper, Fubini, Furlan, and Rossetti (1965), on which the 1966 Fubini paper was based.]

More general local current algebra sum rules, and a survey of earlier work, were also given in R. Dashen and M. Gell-Mann, Representation of Local Current Algebra at Infinite Momentum, *Phys. Rev. Letters* **17**, 340-343 (1966). [19]

Chew, G. F., M. L. Goldberger, F. E. Low, and Y. Nambu (1957). Relativistic Dispersion Relation Approach to Photomeson Production. *Phys. Rev.* **106**, 1345-1355. [5]

Chou, K.-C. (1960). On the Pseudovector Current and Lepton Decays of Baryons and Mesons *J. Exptl. and Theoret. Phys. (U.S.S.R.)* **39**, 703-712 (English translation: *Sov. Phys. JETP* **12**, 492-497 (1961)). [5]

Christenson, J. H., J. W. Cronin, V. L. Fitch, and R. Turlay (1964). Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.* **13**, 138-140. [9]

Conrad, J. M., M. H. Shaevitz, and T. Bolton (1998). Precision Measurements with High-Energy Neutrino Beams. *Rev. Mod. Phys.* **70**, 1341-1392. [18]

Dashen, R. F. (1969). Chiral $SU(3) \otimes SU(3)$ as a Symmetry of the Strong Interactions. *Phys. Rev.* **183**, 1245-1260. [22]

de Forest, T. and J.D. Walecka (1966). Electron Scattering and Nuclear Structure. *Advances in Physics* **15**, 1-109. The general formula for inelastic electron scattering is given on p. 8 of this review. [17]

Faissner, H. et al. (1983). Observation of Neutrino and Antineutrino Induced Coherent Neutral Pion Production off Al^{27}. *Phys. Lett. B* **125**, 230-236. [8]

Feynman, R. P. and M. Gell-Mann (1958). Theory of the Fermi Interaction *Phys. Rev.* **109**, 193-198. [5]

Fubini, S. and G. Furlan (1965). Renormalization Effects for Partially Conserved Currents. *Physics* **1**, 229-247. [14]

Fubini, S., G. Furlan, and S. Rossetti (1965). A Dispersion Theory of Symmetry Breaking. *Nuovo Cimento* **40**, 1171-1193. [21,25]

Fubini, S., Y. Nambu, and V. Wataghin (1958). Dispersion Theory Treatment of Pion Production in Electron-Nucleon Collisions. *Phys. Rev.* **111**, 329-336. [5]

Gell-Mann, M. (1964). The Symmetry Group of Vector and Axial Vector Currents. *Physics* **1**, 63-75. [9,14,17]

Gell-Mann, M. and M. Lévy (1960). The Axial Vector Current in Beta Decay. *Nuovo Cimento* **16**, 705-726. [5,6,10]

Gell-Mann, M., R. J. Oakes, and B. Renner (1968). Behavior of Current Divergences under $SU_3 \times SU_3$. *Phys. Rev.* **175**, 2195-2199. [22]

Goldberger, M. L. and S.B. Treiman (1958). Decay of the Pi Meson. *Phys. Rev.* **110**, 1178-1184. [5,6,11,13]

Goldstone, J., A. Salam, and S. Weinberg (1962). Broken Symetries. *Phys. Rev.* **127**, 965-970. [22]

Gorringe, T. and H. W. Fearing (2004). Induced Pseudoscalar Coupling of the Proton Weak Interaction. *Rev. Mod. Phys.* **76**, 31-91. See in particular Sec. V. [20]

Gottfried, K. (1967). Sum Rule for High-Energy Electron-Proton Scattering. *Phys. Rev. Lett.* **18**, 1174-1177. [19]

Gürsey, F. (1960). On the Symmetries of Strong and Weak Interactions. *Nuovo Cimento* **16**, 230-240. [14]

Joos, P. et al. (1976). Determination of the Nucleon Axial Vector Form-Factor from $\pi\Delta$ Electroproduction Near Threshold. *Phys. Lett. B* **62**, 230-232. [21]

Kataev, A. L. (2004). The Puzzle of the Non-Planar Structure of the QCD Contributions to the Gottfried Sum Rule; arXiv: hep-ph/0412369. [19]

Kitagaki, T. et al. (1986). Charged-Current Exclusive Pion Production in Neutrino-Deuterium Interactions. *Phys. Rev. D* **34**, 2554-2565. [22]

Kopeliovich, B.Z. (2004). PCAC and Shadowing of Low Energy Neutrinos; arXiv:hep-ph/0409079. [8]

Llewellyn Smith, C. H. (1972). Neutrino Reactions at Accelerator Energies. *Physics Reports* **3**, 261-379. For the forward lepton PCAC test and shadowing, see Sec. 2.2B and Sec. 3.8; for the derivation of the neutrino sum rule and its suggestion of "point-like" constituents, see Sec. 2.2C; for the relation between the neutrino sum rule and scaling, see Sec. 3.6B. Llewellyn Smith's review, along with others, was reprinted in *Gauge Theories and Neutrino Physics, Physics Reports Reprint Book Series, Vol. 2* (in memory of Benjamin W. Lee), M. Jacob, ed., (North-Holland, 1978), for which I wrote a short general introduction. [8,19,20]

Low, F. E. (1958). Bremsstrahlung of Very Low-Energy Quanta in Elementary Particle Collisions. *Phys. Rev.* **110**, 974-977. [12,20]

Mangano, M. L. et al. (2001). Physics at the Front-End of a Neutrino Factory: A Quantitative Appraisal; arXiv: hep-ph/0105155. [8]

Nambu, Y. (1959). Discussion remarks, in *Proceedings of the International Conference on High-Energy Physics IX* (1959) (Academy of Science, Moscow, 1960), Vol. 2, pp. 121-122. [14]

Nambu, Y. (1960). Axial Vector Current Conservation in Weak Interactions. *Phys. Rev. Lett.* **4**, 380-382. [5,6]

Nambu, Y. and G. Jona-Lasinio (1961). Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I *Phys. Rev.* **122**, 345-358. [14]

Nambu, Y. and D. Lurié (1962). Chirality Conservation and Soft Pion Production. *Phys. Rev.* **125**, 1429-1436. [6,11-13]

Nambu, Y. and E. Shrauner (1962). Soft Pion Emission Induced by Electromagnetic and Weak Interactions. *Phys. Rev.* **128**, 862-868. [6,10-13,21]

Nishijima, K. (1959). Introduction of a Neutral Pseudoscalar Field and a Possible Connection between Strangeness and Parity. *Nuovo Cimento* **11**, 698-710. [13,14]

Paschos, E. A., M. Sakuda, I. Schienbein, and J. Y. Yu (2004). Comparison of a New Δ Resonance Production Model with Electron and Neutrino Data; arXiv: hep-ph/0408185.

[22]

Pasquini, B., D. Drechsel, and L. Tiator (2005). The Fubini-Furlan-Rossetti Sum Rule Revisited. *Eur. Phys. J. A* **23**, 279-289. [21]

Ravndal, F. (1973). Weak Production of Nuclear Resonances in a Relativistic Quark Model. *Nuovo Cimento A* **18**, 385-415. [8]

Rein, D. and L. M. Sehgal (1981). Neutrino-Excitation of Baryon Resonances and Single Pion Production. *Ann. Phys.* **133**, 79-153. See particularly pp. 117-123. [8,22]

Roper, L. D. (1964). Evidence for a P_{11} Pion-Nucleon Resonance at 556 MeV. *Phys. Rev. Lett.* **12**, 340-342, and private communication. [9]

Rupp, G., F. Kleefeld, and E. van Beveren (2004). Scalar Mesons and Adler Zeros; arXiv: hep-ph/0412078. [10]

Schreiner, P. A. and F. von Hippel (1973a). $\nu p \to \mu^-\Delta^{++}$: Comparison with Theory. *Phys. Rev. Lett.* **30**, 339-342. [22]

Schreiner, P. A. and F. von Hippel (1973b). Neutrino Production of the $\Delta(1236)$. *Nucl. Phys. B* **58**, 333-362. [22]

Schwinger, J. (1958). *Quantum Electrodynamics* (Dover Publications, New York). [4,39]

Sehgal, L. M. (1988). Neutrino Tridents, Conserved Vector Current, and Partially Conserved Axial-Vector Current. *Phys. Rev. D* **38**, 2750-2752. [8]

Shrauner, E. (1963). Chirality Conservation and Soft-Pion Production in Pion-Nucleon Collisions. Phys. Rev. **131**, 1847-1856. [12,13]

Solvay (1968). Fundamental Problems in Elementary Particle Physics (Proceedings of the Fourteenth Conference on Physics at the University of Brussels, October 1967). Interscience, London. My untitled remarks and the ensuing discussion are on pp. 205-214; Chew's comment is on pp. 212-213. In my remarks I attributed the saturation models to Bjorken, but there is no preprint reference; I believe I learned of the models directly from Bjorken when we were both lecturers at the Varenna summer school in July, 1967; see the discussion on page 63 of his lectures published as Bjorken (1968). [19,20]

Tomozawa, Y. (1966). Axial-Vector Coupling Constant Renormalization and the Meson-Baryon Scattering Lengths. *Nuovo Cimento A* **46**, 707-717. [16]

Touschek, B. F. (1957). The Mass of the Neutrino and the Non-Conservation of Parity. *Nuovo Cimento* **5**, 1281-1291. [13]

Weber, A. and L. M. Sehgal (1991). CVC and PCAC in Neutrino-Lepton Interactions. *Nucl. Phys. B* **359**, 262-282. [8]

Weinberg, S. (1966a). Pion Scattering Lengths. *Phys. Rev. Lett.* **17**, 616-621. [16]

Weinberg, S. (1966b). Current Commutator Theory of Multiple Pion Production. *Phys. Rev. Lett.* **16**, 879-883. [16]

Weinberg, S. (1966c). Current-Commutator Calculation of the $K_{\ell 4}$ Form Factors. *Phys. Rev. Lett.* **17**, 336-340. [16]

Weisberger, W. I. (1965). Renormalization of the Weak Axial-Vector Coupling Constant. *Phys. Rev. Lett.* **14**, 1047-1051. [15]

Weisberger, W. I. (1966). Unsubtracted Dispersion Relations and the Renormalization of the Weak Axial-Vector Coupling Constants. *Phys. Rev.* **143**, 1302-1309. [16]

3. Anomalies: Chiral Anomalies and Their Nonrenormalization, Perturbative Corrections to Scaling, and Trace Anomalies to All Orders

Chiral Anomalies and $\pi^0 \to \gamma\gamma$ Decay

I got into the subject of anomalies in an indirect way, through exploration during 1967-1968 of the speculative idea that the muon-electron mass difference could be accounted for by giving the muon an additional magnetic monopole electromagnetic coupling through an axial-vector current, which somehow was nonperturbatively renormalized to zero. After much fruitless study of the integral equations for the axial-vector vertex part, I decided in the spring of 1968 to first try to answer a well-defined question, which was whether the axial-vector vertex in QED was renormalized by multiplication by Z_2, as I had been implicitly assuming. At the time when I turned to this question, I had just started a 6-week visit to the Cavendish Laboratory in Cambridge, England after flying to London with my family on April 21, 1968 (as recorded by my ex-wife Judith in my oldest daughter Jessica's "baby book"). In the Cavendish I shared an office with my former adviser Sam Treiman, and was enjoying the opportunity to try a new project not requiring extensive computer analysis; I had only a month before finished my *Annals of Physics* paper R15 on weak pion production (see Chapter 2), which had required extensive computation, not easy to do in those days when one had to wait hours or even a day for the results of a computer run.

My interest in the multiplicative renormalization question had been piqued by work of van Nieuwenhuizen, in which he had attempted to demonstrate the finiteness to all orders of radiative corrections to μ decay, using an argument based on subtraction of renormalization constants that I knew to be incorrect beyond leading order. I had learned about this work during the previous summer, when I was a lecturer at the Varenna summer school held by Lake Como from July 17-29, 1967, at which van Nieuwenhuizen had given a seminar on this topic that was critiqued by Bjorken, another lecturer. (For further historical details about this, see my review article Adler (2004a) on anomalies and anomaly nonrenormalization, from which much of this commentary has been adapted.) Working in the old Cavendish, I rather rapidly found an inductive multiplicative renormalizability proof, paralleling the one in Bjorken and Drell (1965) for finiteness of Z_2 times the vector vertex. I prepared a detailed outline for a paper describing the proof, but before writing things up, I decided as a check to test whether the formal argument for the closed loop part of the Ward identity worked in the case of the smallest loop diagram. This

is a triangle diagram with one axial and two vector vertices (the AVV triangle; see Fig. 1(a)), which because of Furry's theorem (C invariance) has no analog in the vector vertex case. I knew from a student seminar that I had attended during my graduate study at Princeton that this diagram had been explicitly calculated using a gauge-invariant regularization by Rosenberg (1963), who was interested in the astrophysical process $\gamma_V + \nu \to \gamma + \nu$, with γ_V a virtual photon emitted by a nucleus. I got Rosenberg's paper, tested the Ward identity, and to my astonishment (and Treiman's when I told him the result) found that it failed! I soon found that the problem was that my formal proof used a shift of integration variables inside a linearly divergent integral, which (as I again recalled from student reading) had been analyzed in an Appendix to the classic text of Jauch and Rohrlich (1955), with a calculable constant remainder. For all closed loop contributions to the axial vertex in Abelian electrodynamics with larger numbers of vector vertices (the $AVVVV$, $AVVVVVV$,... loops; see Fig. 1(b)), the fermion loop integrals for fixed photon momenta are highly convergent and the shift of integration variables needed in the Ward identity *is* valid, so proceeding in this fermion loop-wise fashion there were apparently no further additional or "anomalous" contributions to the axial-vector Ward identity. With this fact in the back of my mind I was convinced from the outset that the anomalous contribution to the axial Ward identity would come just from the triangle diagram, with no renormalizations of the anomaly coefficient arising from higher order AVV diagrams with virtual photon insertions.

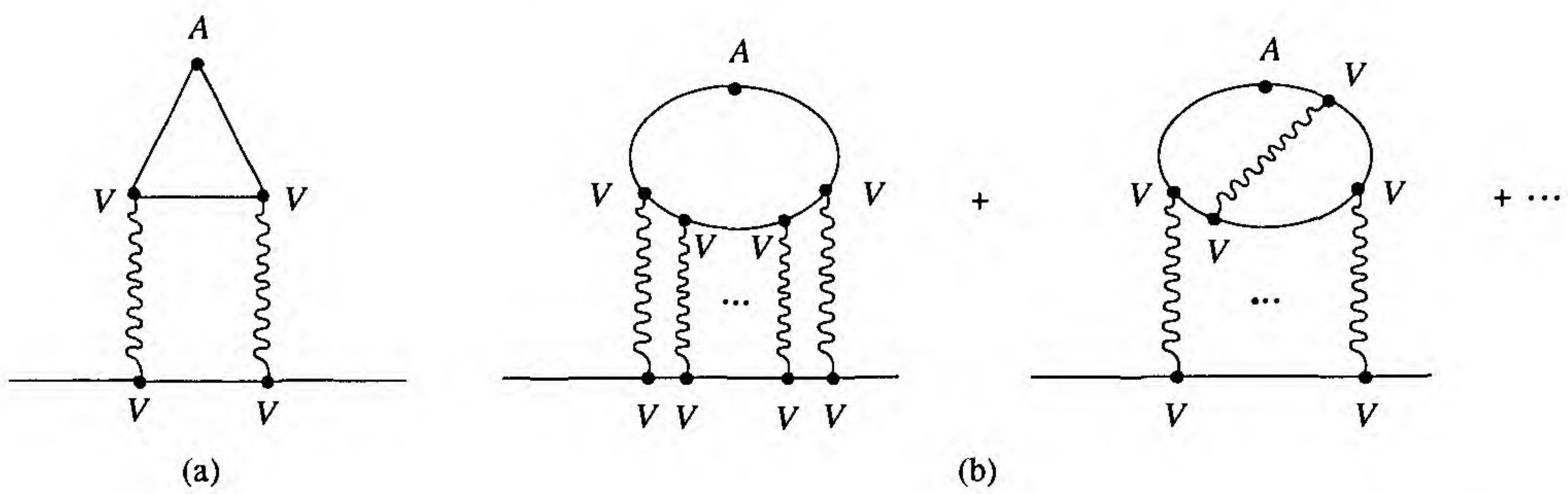


Fig. 1. Fermion loop diagam contributions to the axial-vector vertex part. Solid lines are fermions, and dashed lines are photons. (a) The smallest loop, the AVV triangle diagram. (b) Larger loops with four or more vector vertices, which (when summed over vertex orderings) obey normal Ward identities.

In early June, at the end of my 6 weeks in Cambridge, I returned to the US and then went to Aspen, where I spent the summer working out a manuscript on the properties of the axial anomaly, which became the body (pages 2426-2434) of the final published version (Adler, 1969, R16). Several of the things done there deserve mention, since they were important in later applications. The first was a calculation

of the field theoretic form of the anomaly, giving the now well-known result

$$\partial^\mu j^5_\mu(x) = 2im_0 j^5(x) + \frac{\alpha_0}{4\pi} F^{\xi\sigma}(x) F^{\tau\rho}(x) \epsilon_{\xi\sigma\tau\rho} \quad ,$$

with $j^5_\mu = \overline{\psi}\gamma_\mu\gamma_5\psi$ the axial-vector current (referred to above as A), $j^5 = \overline{\psi}\gamma_5\psi$ the pseudoscalar current, and with m_0 and α_0 the (unrenormalized) fermion mass and coupling constant. The second was a demonstration that because of the anomaly, Z_2 is no longer the multiplicative renormalization constant for the axial-vector vertex, as a result of the diagram drawn in Fig. 1(a) in which the AVV triangle is joined to an electron line with two virtual photons. Instead, the axial-vector vertex is made finite by multiplication by the renormalization constant

$$Z_A = Z_2[1 + \frac{3}{4}(\alpha_0/\pi)^2 \log(\Lambda^2/m^2) + ...] \quad ,$$

thus giving an answer to the question with which I started my investigation. Thirdly, as an application of this result, I showed that the anomaly leads, in fourth order of perturbation theory, to infinite radiative corrections to the current-current theory of $\nu_\mu \mu$ and $\nu_e e$ scattering, but that this infinity can be cancelled between different fermion species by adding appropriate $\nu_\mu e$ and $\nu_e \mu$ scattering terms to the Lagrangian. This result is a forerunner of anomaly cancellation mechanisms in modern gauge theories. It is related to the fact, also discussed in my paper, that the asymptotic behavior of the AVV triangle diagram saturates the bound given by the Weinberg power counting rules, rather than being one power better as is the case for the $AVVVV$ and higher loop diagrams, and has a leading asymptotic term that is a function solely of the external momenta. Finally, I also showed that a gauge invariant chiral generator still exists in the presence of the anomaly. Although not figuring in our subsequent discussion here, in its non-Abelian generalization this was relevant (as reviewed in Coleman, 1989) to later discussions of the $U(1)$ problem in quantum chromodynamics (QCD), leading up to the solution given by 't Hooft (1976).

No sooner was this part of my paper completed than Sidney Coleman arrived in Aspen from Europe, and told me that Bell and Jackiw (published as Bell and Jackiw, 1969) had independently discovered the anomalous behavior of the AVV triangle graph, in the context of a sigma model investigation of the Veltman (1967)–Sutherland (1967) theorem stating that $\pi^0 \to \gamma\gamma$ decay is forbidden in a PCAC calculation. The Sutherland–Veltman theorem is a kinematic statement about the AVV three-point function, which asserts that if the momenta associated with the currents A, V, V are respectively q, k_1, k_2, then the requirement of gauge invariance on the vector currents forces the AVV vertex to be of order qk_1k_2 in the external momenta. Hence when one applies a divergence to the axial-vector vertex and uses the standard PCAC relation (with the quark current $\mathcal{F}^5_{3\mu}$ the analog of $\frac{1}{2}j^5_\mu$)

$$\partial^\mu \mathcal{F}^5_{3\mu}(x) = (f_\pi M_\pi^2/\sqrt{2})\phi_\pi(x) \quad ,$$

with M_π the pion mass, ϕ_π the pion field, and f_π the charged pion decay constant, one finds that the $\pi^0 \to \gamma\gamma$ matrix element is of order $q^2 k_1 k_2$, and hence vanishes in the soft pion limit $q^2 \to 0$. Bell and Jackiw analyzed this result by a perturbative calculation in the σ-model, in which PCAC is formally built in from the outset, and found a non-vanishing result for the $\pi^0 \to \gamma\gamma$ amplitude, which they traced back to the fact that the regularized AVV triangle diagram cannot be defined to satisfy the requirements of both PCAC and gauge invariance. This constituted the "PCAC Puzzle" referred to in the title of their paper. They then proposed to modify the original σ-model by adding further regulator fields with mass-dependent coupling constants in such a manner as to simultaneously enforce gauge invariance and PCAC, thus enforcing the Sutherland–Veltman prediction of a vanishing $\pi^0 \to \gamma\gamma$ decay amplitude. In the words of Bell and Jackiw in their paper, "It has to be insisted that the introduction of this mass dependence of coupling constants is not an arbitrary step in the PCAC context. If a regularization is introduced to define the theory, it must respect any formal properties which are to be appealed to." And again in concluding their paper, they stated "To the complaint that we have changed the theory, we answer that only the revised version embodies simultaneously the ideas of PCAC and gauge invariance."

It was immediately clear to me, in the course of the conversation with Sidney Coleman, that introducing additional regulators to eliminate the anomaly would entail renormalizability problems in σ meson scattering, and was not the correct way to proceed. However, it was also clear that Bell and Jackiw had made an important observation in tying the anomaly to the Sutherland–Veltman theorem for $\pi^0 \to \gamma\gamma$ decay, and that I could use the sigma-model version of the anomaly equation to get a nonzero prediction for the $\pi^0 \to \gamma\gamma$ amplitude, with the whole decay amplitude arising from the anomaly term. I then wrote an Appendix to my paper (pages 2434-2438), clearly delineated from the manuscript that I had finished before Sidney's arrival, in which I gave a detailed rebuttal of the regulator construction, by showing that the anomaly could not be eliminated without spoiling either gauge-invariance or renormalizability. (In later discussions I added unitarity to this list, to exclude the possibility of canceling the anomaly by adding a term to the axial current with a $\partial_\mu/(\partial_\lambda)^2$ singularity.) In this Appendix I also used an anomaly-modified PCAC equation

$$\partial^\mu \mathcal{F}^5_{3\mu}(x) = (f_\pi M_\pi^2/\sqrt{2})\phi_\pi(x) + S\frac{\alpha_0}{4\pi}F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho} \quad ,$$

with S a constant determined by the constituent fermion charges and axial-vector couplings, to obtain a PCAC formula for the $\pi^0 \to \gamma\gamma$ amplitude F^π

$$F^\pi = -(\alpha/\pi)2S\sqrt{2}/f_\pi \quad .$$

Although the axial anomaly, in the context of breakdown of the "pseudoscalar-pseudovector equivalence theorem", had in fact been observed much earlier, start-

ing with Fukuda and Miyamoto (1949) and Steinberger (1949) and continuing to Schwinger (1951), my paper broke new ground by treating the anomaly neither as a baffling calculational result, nor as a field theoretic artifact to be eliminated by a suitable regularization scheme, but instead as a real physical effect (breaking of classical symmetries by the quantization process) with observable physical consequences.

This point of view was not immediately embraced by everyone else. After completing my Appendix I sent Bell and Jackiw copies of my longhand manuscript, and an interesting correspondence ensued. In a letter dated August 25, 1968, Jackiw was skeptical whether one could extract concrete physical predictions from the anomaly, and whether one could augment the divergence of the axial-vector current by a definite extra electromagnetic contribution, as in the modified PCAC equation above. Bell, who was traveling, wrote me on Sept. 2, 1968, and was more appreciative of the possibility of using a modified PCAC to get a formula for the neutral pion decay amplitude, writing "The general idea of adding some quadratic electromagnetic terms to PCAC has been in our minds since Sutherland's η problem. We did not see what to do with it." He also defended the approach he and Jackiw had taken, writing "The reader may be left with the impression that your development is contradictory to ours, rather than complementary. Our first observation is that the σ model interpreted in a conventional way just does not have PCAC. This is already a resolution of the puzzle, and the one which you develop in a very nice way. We, interested in the σ-model only as exemplifying PCAC, choose to modify the conventional procedures, in order to exhibit a model in which general PCAC reasoning could be illustrated in explicit calculation." In recognition of this letter from John Bell, whom I revered, I added a footnote 15 to my manuscript saying "Our results do not contradict those of Bell and Jackiw, but rather complement them. The main point of Bell and Jackiw is that the σ model interpreted in the conventional way, does not satisfy the requirements of PCAC. Bell and Jackiw modify the σ model in such a way as to restore PCAC. We, on the other hand, stay within the conventional σ model, and try to systematize and exploit the PCAC breakdown." This footnote, which contradicts statements made in the text of my paper, has puzzled a number of people; in retrospect, rather than writing it as a paraphrase of Bell's words, I should have quoted directly from Bell's letter.

Following this correspondence, my paper was typed on my return to Princeton and was received by the *Physical Review* on Sept. 24, 1968. (Bell and Jackiw's paper, a CERN preprint dated July 16, 1968, was submitted to *Il Nuovo Cimento*, and received by that journal on Sept. 11, 1968.) My paper was accepted along with a signed referee's report from Bjorken, stating "This paper opens a topic similar to the old controversies on photon mass and nature of vacuum polarization. The lesson there, as I (no doubt foolishly) predict will happen here, is that infinities in diagrams are really troublesome, and that if the cutoff which is used violates a

cherished symmetry of the theory, the results do not respect the symmetry. I will also predict a long chain of papers devoted to the question the author has raised, culminating in a clever renormalizable cutoff which respects chiral symmetry and which, therefore, removes Adler's extra term." Thus, acceptance of the point of view that I had advocated was not immediate, but only followed over time. In 1999, Bjorken was a speaker at my 60th birthday conference at the Institute for Advanced Study, and amused the audience by reading from his report, and then very graciously gave me his file copy, with an appreciative inscription, as a souvenir.

The viewpoint that the anomaly determined the $\pi^0 \to \gamma\gamma$ decay amplitude had significant physical consequences. In the Appendix to my paper, I showed that the value $S = \frac{1}{6}$ implied by the fractionally charged quark model gave a decay amplitude that was roughly a factor of 3 too small. More generally, I showed that a triplet constituent model with charges $(Q, Q-1, Q-1)$ gave $S = Q - \frac{1}{2}$, and so with integrally charged constituents ($Q = 0$ or $Q = 1$) one gets an amplitude that agrees in absolute value, to within the expected accuracy of PCAC, with experiment. I noted in my paper that $Q = 0$, or $S = -\frac{1}{2}$ corresponded to the case in which radiative corrections to weak interactions had been shown to be finite, but this choice for the sign of the $\pi^0 \to \gamma\gamma$ amplitude was soon to be ruled out. Over the next few months Okubo (1969) and Gilman (1969) wrote me letters accompanying preprints which demonstrated, by different methods, that the sign corresponding to a single positive integrally charged constituent going around the triangle loop agrees with experiment. Okubo also analyzed various alternative models for proton constituents, and pointed out that while some are excluded by the experimentally determined value of S, the integrally charged Maki (1964)–Hara (1964) single triplet model (the model that I had considered in my Appendix, but now with $Q = 1$), and the corresponding integrally charged three triplet model of Han and Nambu (1965) (see also Tavkhelidze (1965), Miyamoto (1965), and Nambu (1965)), are both in accord with the empirical value $S \simeq \frac{1}{2}$. In a conference talk a year later, in September 1969 (Adler, 1970a, R17) I reviewed the subject of the anomaly calculation of neutral pion decay, as developed in the papers that had appeared during the preceding year.

The work just described gave the first indications that neutral pion decay provides empirical evidence that can discriminate between different models for hadronic constituents. The correct interpretation of the fact that $S \simeq \frac{1}{2}$ came only later, when what we now call the "color" degree of freedom was introduced in the seminal papers of Bardeen, Fritzsch, and Gell-Mann (1972; reprinted as hep-ph/0211388) and Fritzsch and Gell-Mann (1971/1972; reprinted as hep-ph/0301127). These papers used my calculation of $\pi^0 \to \gamma\gamma$ decay as supporting justification for the tripling of the number of fractionally charged quark degrees of freedom, thus increasing the theoretical value of S for fractionally charged quarks from $\frac{1}{6}$ to $\frac{1}{2}$. The paper of Bardeen, Fritzsch, and Gell-Mann also pointed out that this tripling would show up in a measurement of R, the ratio of hadronic to muon pair production in electron

positron collisions, while noting that "Experiments at present are too low in energy and not accurate enough to test this prediction, but in the next year or two the situation should change.", as indeed it did.

Before leaving the subject of the early history of the anomaly and its antecedents, perhaps this is the appropriate place to mention the paper of Johnson and Low (1966), which showed that the Bjorken (1966)–Johnson–Low (1966) (BJL) method of identifying formal commutators with an infinite energy limit of Feynman diagrams gives, in significant cases, results that differ from the naive field-theoretic evaluation of these commutators. This method was later used by Jackiw and Johnson (1969) and by Boulware and myself (Adler and Boulware, 1969, R18) to show that the AVV axial anomaly can be reinterpreted in terms of anomalous commutators. This line of investigation, however, did not readily lend itself to a determination of anomaly effects beyond leading order. For example, I still have in my files an unpublished manuscript (circa 1966) attempting to use the BJL method to tackle a simpler problem, that of proving that the Schwinger term in quantum electrodynamics (QED) is a c-number to all orders of perturbation theory. I believe that this result is true (and it may well have been proved by now using operator product expansion methods), but I was not able at that time to achieve sufficient control of the BJL limits of high order diagrams with general external legs to give a proof. (See also remarks on this in Chapter 4.)

Anomaly Nonrenormalization

We are now ready to address the issue of the determination of anomalies beyond leading order in perturbation theory. Before the neutral pion low energy theorem could be used as evidence for the charge structure of quarks, one needed to be sure that there were no perturbative corrections to the anomaly and the low energy theorem following from it. As I noted above, the fermion loop-wise argument that I used in my original treatment left me convinced that only the lowest order AVV diagram would contribute to the anomaly, but this was not a proof. This point of view was challenged in the article by Jackiw and Johnson (1969), received by the *Physical Review* on Nov. 25, 1968, who stated "Adler has given an argument to the end that there exist no higher-order effects. He introduced a cutoff, calculated the divergence, and then let the cutoff go to infinity. This is seen in the present context to be equivalent to the second method above. However, we believe that this method may not be reliable because of the dependence on the order of limits." And in their conclusion, they stated "In a definite model the nature of the modification (to the axial-vector current divergence equation) can be determined, but in general only to lowest order in interactions." This controversy with Jackiw and Johnson was the motivation for a more thorough analysis of the nonrenormalization issue undertaken by Bill Bardeen and myself in the fall and winter of 1968-1969 (Adler and Bardeen,

1969, R19) and was cited in the "Acknowledgments" section of our paper, where we thanked "R. Jackiw and K. Johnson for a stimulating controversy which led to the writing of this paper."

The paper with Bardeen approached the problem of nonrenormalization by two different methods. We first gave a general constructive argument for nonrenormalization of the anomaly to all orders, in both quantum electrodynamics and in the σ-model in which PCAC is canonically realized, and we then backed this argument up with an explicit calculation of the leading order radiative corrections to the anomaly, showing that they cancelled among the various contributing Feynman diagrams. The strategy of the general argument was to note that since the anomaly equations written above involve unrenormalized fields, masses, and coupling constants, these equations are well defined only in a cutoff field theory. Thus, for both electrodynamics and the σ-model, we constructed cutoff versions by introducing photon or σ-meson regulator fields with mass Λ. (This was simple for the case of electrodynamics, but more difficult, relying heavily on Bill Bardeen's prior experience with meson field theories, in the case of the σ-model.) In both cases, the cutoff prescription allows the usual renormalization program to be carried out, expressing the unrenormalized quantities in terms of renormalized ones and the cutoff Λ. In the cutoff theories, the fermion loop-wise argument I used in my original anomaly paper is still valid, because regulating boson propagators does not alter the chiral symmetry properties of the theory, and thus it is straightforward to prove the validity of the anomaly equations involving unrenormalized quantities to all orders of perturbation theory.

Taking the vacuum to two γ matrix element of the anomaly equations, and applying the Sutherland–Veltman theorem, which asserts the vanishing of the matrix element of $\partial^\mu j^5_\mu$ at the special kinematic point $q^2 = 0$, Bardeen and I then got exact low energy theorems for the matrix elements $\langle 2\gamma | 2im_0 j^5 | 0\rangle$ (in electrodynamics) and $\langle 2\gamma | (f_\pi M_\pi^2/\sqrt{2})\phi_\pi | 0\rangle$ (in the σ-model) of the "naive" axial-vector divergence at this kinematic point, which were given by the negative of the corresponding matrix element of the anomaly term. However, since we could prove that these matrix elements are finite in the limit as the cutoff Λ approaches infinity, this in turn gave exact low energy theorems for the renormalized, physical matrix elements in both cases. One subtlety that entered into the all orders calculation was the role of photon rescattering diagrams connected to the anomaly term, but using gauge invariance arguments analogous to those involved in the Sutherland–Veltman theorem, we were able to show that these diagrams made a vanishing contribution to the low energy theorem at the special kinematic point $q^2 = 0$. Thus, my paper with Bardeen provided a rigorous underpinning for the use of the $\pi^0 \to \gamma\gamma$ low energy theorem to study the charge structure of quarks.

In our explicit second order calculation, we calculated the leading order radiative corrections to this low energy theorem, arising from addition of a single virtual pho-

ton or virtual σ-meson to the lowest order diagram. We did this by two methods, one involving a direct calculation of the integrals, and the other (devised by Bill Bardeen) using a clever integration by parts argument to bypass the direct calculation. Both methods gave the same answer: the sum of all the radiative corrections is zero, as expected from our general nonrenormalization argument. We also traced the contradictory results obtained in the paper of Jackiw and Johnson to the fact that these authors had studied an axial-vector current (such as $\bar{\psi}\gamma_\mu\gamma_5\psi$ in the σ-model) that is not made finite by the usual renormalizations in the absence of electromagnetism; as a consequence, the naive divergence of this current is not multiplicatively renormalizable. As we noted in our paper, "In other words, the axial-vector current considered by Jackiw and Johnson and its naive divergence are not well-defined objects in the usual renormalized perturbation theory; hence the ambiguous results which these authors have obtained are not too surprising." Our result of a definite, unrenormalized low energy theorem, we noted, came about because "In each model we have studied a *particular* axial-vector current: in spinor electrodynamics, the usual axial-vector current ... and in the σ model the Polkinghorne (1958a,b) axial-vector current ... which, in the absence of electromagnetism, obeys the PCAC condition." It is these axial-vector currents that obey a simple anomaly equation to all orders in perturbation theory, and which give an exact, physically relevant low energy theorem for the naive axial-vector divergence.

This paper with Bill Bardeen should have ended the controversy over whether the anomaly was renormalized, but it didn't. Johnson pointed out in an unpublished report that since the anomaly is mass-independent, it should be possible to calculate it in massless electrodynamics, for which the naive divergence $2im_0j^5$ vanishes and the divergence of the axial-vector current directly gives the anomaly. Moreover, in massless electrodynamics there is no need for mass renormalization, and so if one chooses Landau gauge for the virtual photon propagator, the second order radiative correction calculation becomes entirely ultraviolet finite, with no renormalization counter terms needed. Such a second order calculation was reported by Sen (1970), a Johnson student, who claimed to find nonvanishing second order radiative corrections to the anomaly. However, the calculational scheme proposed by Johnson and used by Sen has the problem that, while ultraviolet finite, there are severe infrared divergences, which if not handled carefully can lead to spurious results. After a long and arduous calculation (Adler, Brown, Wong, and Young, 1971) my collaborators and I were able to show that the zero mass calculation, when properly done, also gives a vanishing second order radiative correction to the anomaly. This confirmed the result I had found with Bardeen, which had by then also been confirmed by different methods in the $m_0 \neq 0$ theory in papers of Abers, Dicus, and Teplitz (1971) and Young, Wong, Gounaris, and Brown (1971).

Even this was not the end of controversies over the nonrenormalization theorem, as discussed in detail in my review Adler (2004a) that focuses specifically on

anomaly nonrenormalization. Suffice it to say here that no objections raised have withstood careful analysis, and there is now a detailed understanding of anomaly nonrenormalization both by perturbative methods, and by non-perturbative methods proceeding from the Callan–Symanzik equations. There is also a detailed understanding of anomaly nonrenormalization in the context of supersymmetric theories, where initial apparent puzzles are now resolved.

Point Splitting Calculations of the Anomaly

At this point let me backtrack, and discuss the role of point-splitting methods in the study of the Abelian electrodynamics anomaly. In the present context, point-splitting was first used in the discussion given by Schwinger (1951) of the pseudoscalar-pseudovector equivalence theorem, to be described in more detail shortly. Almost immediately following circulation of the seminal anomaly preprints in the fall of 1968, Hagen (1969, received Sept. 24, 1968, and a letter to me dated Oct. 16, 1968), Zumino (1969, and a letter to me dated Oct. 7, 1968), and Brandt (1969, received Dec. 17, 1968, and a letter to me dated Oct. 16, 1968) all rederived the anomaly formula by a point-splitting method. Independently, a point-splitting derivation of the anomaly was given by Jackiw and Johnson (1969, received 25 November, 1968), who explicitly made the connection to Schwinger's earlier work (Johnson was a Schwinger student, and was well acquainted with Schwinger's body of work). The point of all of these calculations is that the anomaly can be derived by formal algebraic use of the equations of motion, provided one redefines the singular product $\overline{\psi}(x)\gamma_\mu\gamma_5\psi(x)$ appearing in the axial-vector current by the point-split expression

$$\lim_{x\to x'} \overline{\psi}(x')\gamma_\mu\gamma_5 \exp[-ie\int_{x'}^{x} dx^\lambda B_\lambda]\psi(x) \quad ,$$

and takes the limit $x' \to x$ at the end of the calculation.

Responding to these developments, I appended a "Note added in proof" to my anomaly paper, mentioning the four field-theoretic, point-splitting derivations that had subsequently been given, and adding "Jackiw and Johnson point out that the essential features of the field-theoretic derivation, in the case of external electromagnetic fields, are contained in J. Schwinger, *Phys. Rev.* **82**, 664 (1951)". What to me was an interesting irony emerged from learning of the connection between anomalies and the famous Schwinger (1951) paper on vacuum polarization. I had in fact read Section II and the Appendices of the 1951 paper, when Alfred Goldhaber and I, during our senior year at Harvard (1960-61), did a reading course on quantum electrodynamics with Paul Martin, which focused on papers in Schwinger's reprint volume (Schwinger, 1958). Paul had told us to read the parts of the Schwinger paper that were needed to calculate the VV vacuum polarization loop, but to skip the

rest as being too technical. Reading Section V of Schwinger's paper brought back to mind a brief, forgotten conversation I had had with Jack Steinberger, who was Director of the Varenna Summer School in 1967. Steinberger had told me that he had done a calculation on the pseudovector-pseudoscalar equivalence theorem for $\pi^0 \to \gamma\gamma$, but had gotten different answers in the two cases; also that Schwinger had claimed to reconcile the answers, but that he (Steinberger) couldn't make sense out of Schwinger's argument. Jack had urged me to look at it, which I never did until getting the Jackiw–Johnson preprint, but in retrospect everything fell into place, and the connection to Schwinger's work was apparent.

This now brings me to the question, did Schwinger's paper constitute the discovery of the anomaly? Both Jackiw, in his paper with Johnson, and I were careful to note the connection between Schwinger's (1951) paper and the point-splitting derivations of the anomaly, once it was called to our attention. However, recently some of Schwinger's former students have gone further, arguing that Schwinger was the discoverer of the anomaly and that my paper and that of Bell and Jackiw were merely a "rediscovery" of a previously known result. I believe that this claim goes beyond the published record of what is in Schwinger's paper, as analyzed in detail in Sec. 2.3 and Appendix A of my review Adler (2004a). Stated briefly, Schwinger's calculation was devoted to making the pseudovector calculation give *the same* non-zero answer as the pseudoscalar one, and what Schwinger calls the redefined axial-vector divergence is in fact *not* the divergence of the gauge-invariant axial-vector current, but rather the axial-vector current divergence *minus* the anomaly. In other words, Schwinger's calculation effectively transposes the anomaly term to the left-hand side of the anomaly equation, so that what he evaluates is the effective Lagrangian arising from the left-hand side of the equation

$$\partial^\mu j^5_\mu(x) - \frac{\alpha_0}{4\pi} F^{\xi\sigma}(x) F^{\tau\rho}(x) \epsilon_{\xi\sigma\tau\rho} = 2im_0 j^5(x) \quad ,$$

which then necessarily gives the same result as calculation of an effective Lagrangian from the right-hand side, which is pseudoscalar coupling. There is no gauge-invariant axial-vector current for which the combination on the left-hand side is the divergence, but as shown in Eqs. (58) and (59) of R16, there is a gauge-non-invariant axial-vector current which has this divergence.

The use of a point-splitting method was of course important and fruitful, and in retrospect, the axial anomaly is hidden within Schwinger's calculation. But Schwinger never took the crucial step of observing that the axial-vector current matrix elements cannot, in a renormalizable quantum theory, be made to satisfy all of the expected classical symmetries. And more specifically, he never took the step of defining a gauge-invariant axial-vector current by point splitting, which has a well-defined anomaly term in its divergence, with the anomaly term completely accounting for the disagreement between the pseudoscalar and pseudovector calculations of neutral pion decay. So I would say that although Schwinger took steps in

the right direction, particularly in noting the utility of point-splitting in defining the axial-vector current, his 1951 paper *obscured* the true physics and does not mark the discovery of the anomaly. This happened only much later, in 1968, and led to a flurry of activity by many people. My view is supported, I believe, by the fact that Schwinger's calculation seemed arcane, even to people (like Steinberger) with whom he had talked about it and to colleagues familiar with his work, and exerted no influence on the field until after preprints on the seminal work of 1968 had appeared.

The Non-Abelian Anomaly, Its Nonrenormalization and Geometric Interpretation

Since in the chiral limit the AVV triangle is identical to an AAA triangle (as is easily seen by an argument involving anticommutation of a γ_5 around the loop), I knew already in unpublished notes dating from the late summer of 1968 that the AAA triangle would also have an anomaly; a similar observation was also made by Gerstein and Jackiw (1969). From fragmentary calculations begun in Aspen I suspected that higher loop diagrams might have anomalies as well, so after the nonrenormalization work with Bill Bardeen was finished I suggested to Bill that he work out the general anomaly for larger diagrams. (I was at that point involved in other calculations with Wu-Ki Tung, on the perturbative breakdown of scaling formulas such as the Callan–Gross relation, to be discussed shortly.) I showed Bill my notes, which turned out to be of little use, but which contained a very pertinent remark by Roger Dashen that including charge structure (which I had not) would allow a larger class of potentially anomalous diagrams. Within a few weeks, Bill carried out a brilliant calculation, by point-splitting methods, of the general anomaly in both the Abelian *and* the non-Abelian cases (Bardeen, 1969). Expressed in terms of vector and axial-vector Yang–Mills field strengths

$$F_V^{\mu\nu}(x) = \partial^\mu V^\nu(x) - \partial^\nu V^\mu(x) - i[V^\mu(x), V^\nu(x)] - i[A^\mu(x), A^\nu(x)] \quad ,$$

$$F_A^{\mu\nu}(x) = \partial^\mu A^\nu(x) - \partial^\nu A^\mu(x) - i[V^\mu(x), A^\nu(x)] - i[A^\mu(x), V^\nu(x)] \quad ,$$

Bardeen's result takes the form

$$\begin{aligned}\partial^\mu J_{5\mu}^\alpha(x) = {} & \text{normal divergence term} \\ & + (1/4\pi^2)\epsilon_{\mu\nu\sigma\tau}\mathrm{tr}_I[\lambda_A^\alpha[(1/4)F_V^{\mu\nu}(x)F_V^{\sigma\tau}(x) + (1/12)F_A^{\mu\nu}(x)F_A^{\sigma\tau}(x) \\ & + (2/3)iA^\mu(x)A^\nu(x)F_V^{\sigma\tau}(x) + (2/3)iF_V^{\mu\nu}(x)A^\sigma(x)A^\tau(x) \\ & + (2/3)iA^\mu(x)F_V^{\nu\sigma}(x)A^\tau(x) - (8/3)A^\mu(x)A^\nu(x)A^\sigma(x)A^\tau(x)] \quad ,\end{aligned}$$

with tr_I denoting a trace over internal degrees of freedom, and λ_A^α the internal symmetry matrix associated with the axial-vector external field. In the Abelian case,

with trivial internal symmetry structure, the terms involving two or three factors of $A^{\mu,\nu,\cdots}$ vanish by antisymmetry of $\epsilon_{\mu\nu\sigma\tau}$, and there are only AVV and AAA triangle anomalies. When there is non-trivial internal symmetry or charge structure, there are anomalies associated with the box and pentagon diagrams as well, confirming Dashen's intuition mentioned earlier. Bardeen notes that whereas the triangle and box anomalies result from linear divergences associated with these diagrams, the pentagon anomalies arise not from linear divergences, but rather from the definition of the box diagrams to have the correct vector current Ward identities. Bardeen also notes, in his conclusion, another prophetic remark of Dashen, to the effect that the pentagon anomalies should add anomalous terms to the PCAC low energy theorems for five pion scattering; I shall return to this shortly.

There are two distinct lines of argument leading to the conclusion that the non-Abelian chiral anomaly also has a nonrenormalization theorem, and is given exactly by Bardeen's leading order calculation. The first route parallels that used in the Abelian case, involving variously a loop-wise regulator construction, explicit fourth order calculation, and an argument using the Callan–Symanzik equations; for detailed references, see Adler (2004a). The conclusion in all cases is that the Adler–Bardeen theorem extends to the non-Abelian case. Heuristically, what is happening is that except for a few small one-fermion loop diagrams, non-Abelian theories, just like Abelian ones, are made finite by gauge invariant regularization of the gluon propagators. But this regularization has no effect on the chiral properties of the theory, and therefore does not change its anomaly structure, which can thus be deduced from the structure of the few small fermion loop diagrams for which naive classical manipulations break down.

The second route leading to the conclusion that the non-Abelian anomaly is nonrenormalized might be termed "algebraic/geometrical", and consists of two steps. The first step consists of a demonstration that the higher order terms in Bardeen's non-Abelian formula are completely determined by the leading, Abelian anomaly. During a summer visit to Fermilab in 1971, I collaborated with Ben Lee, Sam Treiman, and Tony Zee (Adler, Lee, Treiman, and Zee 1971, R20) in a calculation of a low energy theorem for the reaction $\gamma+\gamma\to\pi+\pi+\pi$ in both the neutral and charged pion cases. This was motivated in part by discrepancies in calculations that had just appeared in the literature, and in part by its relevance to theoretical unitarity calculations of a lower bound on the $K_L^0\to\mu^+\mu^-$ decay rate. Using PCAC, we showed that the fact that the $\gamma+\gamma\to 3\pi$ matrix elements vanish in the limit when a final π^0 becomes soft, together with photon gauge invariance, relates these amplitudes to the matrix elements F^π for $\gamma+\gamma\to\pi^0$ and $F^{3\pi}$ for $\gamma\to\pi^0+\pi^++\pi^-$, and moreover, gives a relation between the latter two matrix elements,

$$eF^{3\pi}=f^{-2}F^{\pi}\quad,\ f=\frac{f_\pi}{\sqrt{2}}\quad .$$

Thus all of the matrix elements in question are uniquely determined by F^{π}, which itself is determined by the AVV anomaly calculation. An identical result for the same reactions was independently given by Terent'ev (Terentiev) (1971). In the meantime, in a beautiful formal analysis, Wess and Zumino (1971) showed that the current algebra satisfied by the flavor $SU(3)$ octet of vector and axial-vector currents implies integrability or "consistency" conditions on the non-Abelian axial-vector anomaly, which are satisfied by the Bardeen formula, and conversely, that these constraints uniquely imply the Bardeen structure up to an overall factor, which is determined by the Abelian AVV anomaly. By introducing an auxiliary pseudoscalar field, Wess and Zumino were able to write down a local action obeying the anomalous Ward identities and the consistency conditions. (There is no corresponding local action involving just the vector and axial-vector currents, since if there were, the anomalies could be eliminated by a local counterterm.) Wess and Zumino also gave expressions for the processes $\gamma \to 3\pi$ and $2\gamma \to 3\pi$ discussed by Adler et al. and Terentiev, as well as giving the anomaly contribution to the five pseudoscalar vertex. The net result of these three simultaneous pieces of work was to show that the Bardeen formula has a rigidly constrained structure, up to an overall factor given by the $\pi^0 \to \gamma\gamma$ decay amplitude.

The second step in the "algebraic/geometric" route to anomaly renormalization is a celebrated paper of Witten (1983), which shows that the Wess–Zumino action has a representation as the integral of a fifth rank antisymmetric tensor (constructed from the auxiliary pseudoscalar field) over a five-dimensional disk of which four-dimensional space is the boundary. In addition to giving a new interpretation of the Wess–Zumino action Γ, Witten's argument also gave a constraint on the overall factor in Γ that was not determined by the Wess–Zumino consistency argument. Witten observed that his construction is not unique, because a closed five-sphere intersecting a hyperplane gives two ways of bounding the four-sphere along the equator with a five dimensional hemispherical disk. Requiring these two constructions to give the same value for $\exp(i\Gamma)$, which is the way the anomaly enters into a Feynman path integral, requires integer quantization of the overall coefficient in the Wess–Zumino–Witten action. This integer can be read off from the AVV triangle diagram, and for the case of an underlying color $SU(N_c)$ gauge theory turns out to be just N_c, the number of colors.

To summarize, the "algebraic/geometric" approach shows that the Bardeen anomaly has a unique structure, up to an overall constant, and moreover that this overall constant is constrained by an integer quantization condition. Hence once the overall constant is fixed by comparison with leading order perturbation theory (say in QCD), it is clear that this result must be exact to all orders, since the presence of renormalizations in higher orders of the strong coupling constant would lead to violations of the quantization condition.

The fact that non-Abelian anomalies are given by an overall rigid structure

has important implications for quantum field theory. For example, the presence of anomalies spoils the renormalizability of non-Abelian gauge theories and requires the cancellation of gauged anomalies between different fermion species (see Gross and Jackiw (1972), Bouchiat, Iliopoulos, and Meyer (1972), and Weinberg (1973)), through imposition of the condition $\mathrm{tr}\{T_\alpha, T_\beta\}T_\gamma = 0$ for all α, β, γ, with T_α the coupling matrices of gauge bosons to left-handed fermions. The fact that anomalies have a rigid structure then implies that once these anomaly cancellation conditions are imposed for the lowest order anomalous triangle diagrams, no further conditions arise from anomalous square or pentagon diagrams, or from radiative corrections to these leading fermion loop diagrams. Other places where the one-loop geometric structure of non-Abelian anomalies enters are in instanton physics, and in the 't Hooft anomaly matching conditions. These and other chiral anomaly applications are discussed in more detail in my review Adler (2004a), and also in my *Encyclopedia of Mathematical Physics* article Adler (2004b). Both of these sources give extensive references to recent review articles and books on anomalies, which update the 1970 reviews given in my Brandeis lectures (Adler, 1970b) and in Jackiw's Brookhaven lectures (Jackiw, 1970).

Perturbative Corrections to Scaling

While finishing the paper with Bardeen on anomaly nonrenormalization, I had embarked on a different set of perturbative calculations with Wu-Ki Tung; these became a forerunner of a different kind of "anomaly", the anomalous scaling observed in deep inelastic electron and neutrino scattering. Our starting point was the question of whether applications of the Bjorken (1966) limit technique, which assumed that the asymptotic behavior of time-ordered products is given by the "naive" or free field theory equal time commutator, would be modified in perturbation theory. Strong hints in this direction had been given in a paper of Johnson and Low (1966), which showed that the "Bjorken–Johnson–Low" limit can produce anomalous commutators, and related results were also obtained in an earlier paper of Vainshtein and Ioffe (1967); our aim was to do calculations focusing on several physically important applications not covered in this previous work. These were the calculation by Bjorken (1966) of the radiative corrections to β-decay, the Bjorken (1967) backward-neutrino-scattering asymptotic sum rule, and the Callan–Gross (1969) relation relating the ratio of the longitudinal to transverse deep inelastic electron scattering cross sections to the constitution of the electric current, with the latter an application both of the Bjorken–Johnson–Low limit method, and of the later proposal by Bjorken (1969) of scaling of the deep inelastic structure functions.

For our test model, we considered an $SU(3)$ triplet of spin-1/2 particles bound by exchange of a massive singlet gluon, which we took as either a vector, scalar, or pseudoscalar. The results of the vector exchange calculation, to leading order of

perturbation theory, were reported in Adler and Tung (1969), R21, while additional leading order results in the scalar and pseudoscalar gluon cases, and some fourth order results, were given in the follow-up paper Adler and Tung (1970), R22. We concluded that the Callan–Gross relation for spin-1/2 quarks, which asserts the vanishing of $q^2\sigma_L(q^2,\omega)$ for large q^2 with fixed scaling variable ω, breaks down in leading order of perturbation theory. A similar conclusion was also reached by Jackiw and Preparata (1969a,b), whose first paper appears in the same issue of *Physical Review Letters* as our paper R21. Tung and I related the breakdown of the Callan–Gross relation to a corresponding breakdown of Bjorken's backward neutrino sum rule. We also showed that the certain current commutators receive a systematic pattern of logarithmic asymptotic corrections, and calculated the leading perturbative correction to the logarithmically divergent part of the radiative corrections to β decay. Tung (1969), while still at the Institute, and Jackiw and Preparata (1969c), went on to carry out general analyses of the range of validity and breakdown of the Bjorken–Johnson–Low limit in perturbation theory.

These papers had a number of implications for subsequent developments. The logarithmic deviations from the Callan–Gross relation were soon understood in a more systematic way through the Wilson (1969) operator product expansion and the Callan (1970)–Symanzik (1970) equations, which gave anomalous dimensions in accord with the leading order results obtained by Tung and me and by Jackiw and Preparata, and with the fourth order results obtained by Tung and me in R22; for a discussion of this, see Bég (1975). The fact that perturbative field theory gives strong violations of scaling led to a skepticism as to whether field theory could describe the strong interactions at all. For example, Fritzsch and Gell-Mann (1971/1972), in their long paper on "Light Cone Current Algebra", remarked that "The renormalized perturbation theory, taken term by term, reveals various pathologies in commutators of currents. Not only are there in each order logarithmic singularities on the light cone, which destroy scaling, and violations of the rule that $\sigma_L/\sigma_T \to 0$ in the Bjorken limit, but also a careful perturbation theory treatment show the existence of higher singularities on the light cone..." This was one of their motivations for introducing the light cone algebra, which abstracted from field theory algebraic relations that led to scaling and parton model results, with the field theory itself being discarded.

At the same time, there were also thoughts that a renormalization group fixed point in field theory might provide a remedy. In the same article, Fritzsch and Gell-Mann noted that in the context of a singlet vector gluon theory, "we must imagine that the sum of perturbation theory yields the special case of a 'finite vector theory'[27][reference to Gell-Mann and Low, and Baker and Johnson] if we are to bring the vector gluon theory and the basic algebra into harmony." Quite independently, in a conference talk at Princeton that I gave in October of 1971 (published considerably later as Adler (1974), R23), in Section 2.4, on "Questions raised by the breakdown of the BJL limit", I made the remark "Can one make a

consistent calculational scheme in which Bjorken limits, the Callan–Gross relation and scaling are all valid? This is a *real challenge* to theorists...Perhaps a successful approach would involve summation of perturbation theory graphs plus use of the Gell-Mann–Low eigenvalue condition (see sect. 3)." (I made these comments at just the time when I was working on a possible eigenvalue condition in quantum electrodynamics, growing out of the work of Gell-Mann and Low, and Johnson, Baker, and Willey, as described below in Chapter 4. The relevance of an eigenvalue to power law behavior was also pointed out in the papers of Callan (1972) and of Christ, Hasslacher, and Mueller (1972), which I included as references when I edited my 1971 conference talk in the fall of 1972.) However, in the field theories then under consideration, there was an obstacle to realizing this idea. As I noted in Sec. 3 of my Princeton talk, for singlet gluon theories the renormalization group methods suggested either no simple scaling behavior (if there were no renormalization group fixed point at which the β function had a zero), or power law deviations from scaling of the form $(q^2)^{-\gamma}$ (if there were a fixed point at a nonzero coupling value λ_0 where β vanished, with γ the value of the anomalous dimension at the fixed point). Since in a strong coupling theory γ would be expected to be large at the fixed point, power law deviations from scaling looked to be too large to agree with experiment.

It took another eighteen months for this obstacle to be overcome. Three developments were involved: the introduction of the modern form of "color" as a tripling of the fractionally charged quark degrees of freedom by Bardeen, Fritzsch, and Gell-Mann (1972), the non-Abelian gauging of this form of color by Fritzsch and Gell-Mann (1972), and finally, in line with Gell-Mann's dictum "Nature reads the books of free field theory", a search for field theories that would have almost free behavior in the scaling limit. The conclusion of this search, the discovery of the asymptotic freedom of non-Abelian gauge theories and its implications by Gross, Politzer, and Wilczek, in the end proved a realization of the field-theoretic route that been contemplated by various people in 1971. In asymptotically free theories, because the renormalization group fixed point (the Gell-Mann–Low eigenvalue) is at zero coupling, where the anomalous dimension γ vanishes, the deviations from scaling are not powers of q^2, but rather only powers of $\log q^2$, with exponents that can be calculated in leading order of perturbation theory. Thus the deviations from scaling predicted by non-Abelian gauge theories, and specifically by quantum chromodynamics (QCD) as the theory of the strong interactions, are much weaker than would be expected for singlet gluon theories, and are compatible with experiment.

Returning briefly to the calculations that Tung and I did, our results for the radiative corrections to β-decay in the singlet vector gluon model turned out later to have applications in the QCD context. They can be converted to the realistic case of the octet gluon of QCD by multiplication by a color factor, as discussed in the review of Sirlin (1978), and so have become part of the technology for calculating radiative corrections to weak processes.

Trace Anomalies to All Orders

In an influential paper Wilson (1969) proposed the operator product expansion, incorporating ideas on the approximate scale invariance of the strong interactions suggested by Mack (1968). As one of the applications of his technique, Wilson discussed $\pi^0 \to 2\gamma$ decay and the axial-vector anomaly from the viewpoint of the short distance singularity of the coordinate space AVV three-point function. Using these methods, Crewther (1972) and Chanowitz and Ellis (1972) investigated the short distance structure of the three-point function $\theta V_\mu V_\nu$, with $\theta = \theta^\mu_\mu$ the trace of the energy-momentum tensor, and concluded that this is also anomalous, thus confirming earlier indications of a perturbative trace anomaly obtained in a study of broken scale invariance by Coleman and Jackiw (1971). Letting $\Delta_{\mu\nu}(p)$ be the momentum space $\theta V_\mu V_\nu$ three point function, and $\Pi_{\mu\nu}$ be the corresponding $V_\mu V_\nu$ two-point function, the naive Ward identity $\Delta_{\mu\nu}(p) = (2 - p_\sigma \partial/\partial p_\sigma)\Pi_{\mu\nu}(p)$ is modified to

$$\Delta_{\mu\nu}(p) = \left(2 - p_\sigma \frac{\partial}{\partial p_\sigma}\right)\Pi_{\mu\nu}(p) - \frac{R}{6\pi^2}(p_\mu p_\nu - \eta_{\mu\nu}p^2) \quad ,$$

with the trace anomaly coefficient R given by

$$R = \sum_{i,\mathrm{spin}\frac{1}{2}} Q_i^2 + \frac{1}{4}\sum_{i,\mathrm{spin}\,0} Q_i^2 \quad .$$

Thus, for QED, with a single fermion of charge e, the anomaly term is $-[2\alpha/(3\pi)](p_\mu p_\nu - \eta_{\mu\nu}p^2)$. In a subsequent paper, Chanowitz and Ellis (1973) showed that the fourth order trace anomaly can be read off directly from the coefficient of the leading logarithm in the asymptotic behavior of $\Pi_{\mu\nu}(p)$, giving to next order an anomaly coefficient $-2\alpha/(3\pi) - \alpha^2/(2\pi^2)$. Thus, their fourth order argument indicated a direct connection between the trace anomaly and the renormalization group β function.

My involvement with trace anomalies began roughly five years later, when *Physical Review* sent me for refereeing a paper by Iwasaki (1977). In this paper, which noted the relevance to trace anomalies, Iwasaki proved a kinematic theorem on the vacuum to two photon matrix element of the trace of the energy-momentum tensor, that is an analog of the Sutherland–Veltman theorem for the vacuum to two photon matrix element of the divergence of the axial-vector current. Just as the latter has a kinematic zero at $q^2 = 0$, Iwasaki showed that the kinematic structure of the vacuum to two photon matrix element of the energy-momentum tensor implies, when one takes the trace, that there is also a kinematic zero at $q^2 = 0$, irrespective of the presence of anomalies (just as the Sutherland–Veltman result holds in the presence of anomalies). Reading this article suggested the idea that just as the Sutherland–Veltman theorem can be used as part of an argument to prove nonrenormalization of the axial-vector anomaly, Iwasaki's theorem could be used to analogously calculate the trace anomaly to all orders. (In addition to writing a

favorable report on Iwasaki's paper, I invited him to spend a year at the IAS, which he did during the 1977-78 academic year.) During the spring of 1976 I wrote an initial preprint attempting an all orders calculation of the trace anomaly in quantum electrodynamics, but this had an error pointed out to me by Baqi Bég. Over the summer of 1976 I then collaborated with two local postdocs, John Collins (at Princeton) and Anthony Duncan (at the Institute), to work out a corrected version (Adler, Collins, and Duncan, 1977, R24). Collins and Duncan simultaneously teamed up with another Institute postdoc, Satish Joglekar, to apply similar ideas to quantum chromodynamics, published as Collins, Duncan, and Joglekar (1977), and independently the same result for QCD was obtained by N. K. Nielsen (1977). Similar results were given in a preprint of Minkowski (1976), which grew out of discussions in the Gell-Mann group at Cal Tech in which the role of the β function in the trace anomaly formula, and its implications for generating the scale of the strong interactions, were appreciated (C. T. Hill, private communication, 2005, and P. Minkowski, private communication, 2005).

In the simpler case of QED, the argument based on Iwasaki's theorem is given in Section II of R24. The basic idea is to use Iwasaki's result for the vacuum to two photon matrix element of the trace of the energy momentum tensor, together with expressions for the electron to electron and the vacuum to two photon matrix elements of the "naive" trace $m_0\overline{\psi}\psi$ given by application of the Callan–Symanzik equations. The final result for the trace is given by

$$\theta^{\mu}_{\mu} = [1+\delta(\alpha)]m_0\overline{\psi}\psi + \frac{1}{4}\beta(\alpha)N[F_{\lambda\sigma}F^{\lambda\sigma}] + \ldots \quad ,$$

with $N[\ \ ]$ an explicitly defined subtracted operator, with ... indicating terms that vanish by the equations of motion, and with $\delta(\alpha)$ and $\beta(\alpha)$ the renormalization group functions defined by $1+\delta(\alpha) = (m/m_0)\partial m_0/\partial m$ and $\beta(\alpha) = (m/\alpha)\partial\alpha/\partial m$. The first two terms in the power series expansion of the coefficient of the $F_{\lambda\sigma}F^{\lambda\sigma}$ term in the trace agree with the fourth-order calculation of Chanowitz and Ellis. The trace equation in QCD has a similar structure, again with the β function appearing as the anomaly coefficient. The fact that the trace anomaly coefficient is given by the appropriate β function extends to the supersymmetric case, and leads to interesting issues that are reviewed in the final section of Adler (2004a). The appearance of the β function in the anomaly coefficient has also played a role in the inference of the structure of effective Lagrangians from the form of the trace anomaly; see, for example, Pagels and Tomboulis (1978) for an application to QCD, and Veneziano and Yankielowicz (1982) for an application to supersymmetric Yang–Mills theory.

References for Chapter 3

Abers, E. S., D. A. Dicus, and V. L. Teplitz (1971). Ward Identities for η Decay in Perturbation Theory. *Phys. Rev. D* **3**, 485-497. [38]

Adler, S. L. (1969) R16. Axial-Vector Vertex in Spinor Electrodynamics. *Phys. Rev.* **177**, 2426-2438. [31-35,40]

Adler, S. L. (1970a) R17. π^0 Decay, in *High-Energy Physics and Nuclear Structure*, S. Devons, ed. (Plenum Press, New York), pp. 647-655. [35]

Adler, S. L. (1970b). Perturbation Theory Anomalies, in *Lectures on Elementary Particles and Quantum Field Theory*, Vol. 1, S. Deser, M. Grisaru, and H. Pendleton, eds. (M.I.T. Press, Cambridge, MA), pp. 3-164. [44]

Adler, S. L. (1974) R23. Anomalies in Ward Identities and Current Commutation Relations, in *Local Currents and Their Applications, Proceedings of an Informal Conference*, Princeton, NJ, 8-10 October, 1971, D. H. Sharp and A. S. Wightman, eds. (North-Holland, Amsterdam and American Elsevier, New York), pp. 142-168. Pages 162-168 are reprinted here. Although overtaken by subsequent events, this proceedings was nonetheless seen through to publication by the editors, who in their Preface noted that "Since the speakers chose to survey the achievements and prospects of their subjects, they concentrated on essentials. As a result, this collection of their talks seems to the organizers to provide a very useful review of the state of the art as of 1971." The published version was based on a handwritten manuscript prepared from my notes by David Sharp, which he sent me with a letter dated September 20, 1972, asking me to check it and add references. That is why the references include papers that were circulated in 1972, after the conference took place. [45,46,59]

Adler, S. L. (2004a). Anomalies to All Orders; arXiv: hep-th/0405040. Published in *Fifty Years of Yang-Mills Theory*, G. 't Hooft, ed. (World Scientific, Singapore, 2005). [30,38,40,42,44,48]

Adler, S. L. (2004b). Anomalies; arXiv: hep-th/0411038. To appear in *Encyclopedia of Mathematical Physics*, to be published by Elsevier in early 2006. [44]

Adler, S. L. and W. A. Bardeen (1969) R19. Absence of Higher-Order Corrections in the Anomalous Axial-Vector Divergence Equation. *Phys. Rev.* **182**, 1517-1536. [36-38,42]

Adler, S. L. and D. G. Boulware (1969) R18. Anomalous Commutators and the Triangle Diagram. *Phys. Rev.* **184**, 1740-1744. [36]

Adler, S. L., R. W. Brown, T. F. Wong, and B.-L. Young (1971). Vanishing of the Second-Order Correction to the Triangle Anomaly in Landau-Gauge, Zero-Fermion-Mass Quantum Electrodynamics. *Phys. Rev. D* **4**, 1787-1808. [38]

Adler, S. L., J. C. Collins, and A. Duncan (1977) R24. Energy-Momentum-Tensor Trace Anomaly in Spin-1/2 Quantum Electrodynamics. *Phys. Rev. D* **15**, 1712-1721. [48]

Adler, S. L., B. W. Lee, S. B. Treiman, and A. Zee (1971) R20. Low Energy Theorem for

$\gamma + \gamma \to \pi + \pi + \pi$. *Phys. Rev. D* **4**, 3497-3501. [42,43,65,66]

Adler, S. L. and W.-K. Tung (1969) R21. Breakdown of Asymptotic Sum Rules in Perturbation Theory. *Phys. Rev. Lett.* **22**, 978-981. [44-46]

Adler, S. L. and W.-K. Tung (1970) R22. Bjorken Limit in Perturbation Theory. *Phys. Rev. D* **1**, 2846-2859. [44-46]

Bardeen, W. (1969). Anomalous Ward Identities in Spinor Field Theories. *Phys. Rev.* **184**, 1848-1859. [41-43]

Bardeen, W. A., H. Fritzsch, and M. Gell-Mann (1972). Light Cone Current Algebra, π^0 Decay, and e^+e^- Annihilation, CERN preprint TH. 1538, 21 July, 1972. Later published in *Scale and Conformal Symmetry in Hadron Physics*, R. Gatto, ed. (Wiley, New York, 1973), pp. 139-151; reissued recently as arXiv: hep-ph/0211388. See also R. J. Oakes, Summary–Second Week, *Acta Phys. Austriaca Suppl.* **IX**, 905-909, which summarizes remarks by Gell-Mann at the Schladming Winter School, Feb. 21-Mar. 4, 1972. [35,46]

Bég, M. A. B. (1975). Anomalous Algebras and Neutrino Sum Rules. *Phys. Rev. D* **11**, 1165-1170. [45]

Bell, J. S. and R. Jackiw (1969). A PCAC Puzzle: $\pi^0 \to \gamma\gamma$ in the σ-Model. *Nuovo Cimento A* **60**, 47-61. [32-34,40]

Bjorken, J. D. (1966). Applications of the Chiral $U(6) \otimes U(6)$ Algebra of Current Densities. *Phys. Rev.* **148**, 1467-1478. [36,44-46]

Bjorken, J. D. (1967). Inequality for Backward Electron- and Muon-Nucleon Scattering at High Momentum Transfer. *Phys. Rev.* **163**, 1767-1769. [44,45]

Bjorken, J. D. (1969). Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.* **179**, 1547-1553. [44-46]

Bjorken, J. D. and S. D. Drell (1965). *Relativistic Quantum Fields* (McGraw-Hill, New York). [30]

Bouchiat, C., J. Iliopoulos, and Ph. Meyer (1972). An Anomaly-Free Version of Weinberg's Model. *Phys. Lett. B* **38**, 519-523. [44]

Brandt, R. A. (1969). Axial-Vector Current in Spinor Electrodynamics. *Phys. Rev.* **180**, 1490-1502. [39]

Callan, C. G. (1970). Broken Scale Invariance in Scalar Field Theory. *Phys. Rev. D* **2**, 1541-1547. [39,42,45,48,55,58,59]

Callan, C. G. (1972). Broken Scale Invariance and Asymptotic Behavior. *Phys. Rev. D* **5**, 3202-3210. [46]

Callan, C. G. and D. J. Gross (1969). High-Energy Electroproduction and the Constitution of the Electric Current. *Phys. Rev. Lett.* **22**, 156-159. [44-46]

Chanowitz, M. S. and J. Ellis (1972). Canonical Anomalies and Broken Scale Invariance. *Phys. Lett. B* **40**, 397-400. [47]

Chanowitz, M. S. and J. Ellis (1973). Canonical Trace Anomalies. *Phys. Rev. D* **7**, 2490-2506. [47,48]

Christ, N., B. Hasslacher, and A. H. Mueller (1972). Light-Cone Behavior of Perturbation Theory. *Phys. Rev. D* **6**, 3543-3562. [46]

Coleman, S. (1989). *Aspects of Symmetry* (Cambridge University Press, Cambridge), pp. 307-327. [32]

Coleman, S. and R. Jackiw (1971). Why Dilatation Generators do not Generate Dilatations. *Ann. Phys.* **67**, 552-598. [47]

Collins, J. C., A. Duncan, and S. D. Joglekar (1977). Trace and Dilatation Anomalies in Gauge Theories. *Phys. Rev. D* **16**, 438-449. [48]

Crewther, R. J. (1972). Nonperturbative Evaluation of the Anomalies in Low-Energy Theorems. *Phys. Rev. Lett.* **28**, 1421-1424. [47]

Fritzsch, H. and M. Gell-Mann (1971/1972). Light Cone Algebra, talk at the 1971 Coral Gables Conference, extended into a preprint a few months later; reissued recently as arXiv: hep-ph/0301127. [35,45]

Fritzsch, H. and M. Gell-Mann (1972). Current Algebra: Quarks and What Else?, in *Proceedings of the XVI International Conference on High Energy Physics*, Chicago–Batavia, IL, J. D. Jackson and A. Roberts, eds. (National Accelerator Laboratory, Batavia), Vol. 2, pp. 135-165. See page 140 for a discussion of the gauge theory of color octet gluons. [46]

Fukuda, H. and Y. Miyamoto (1949). On the γ-Decay of Neutral Meson. *Progr. Theor. Phys.* **4**, 235 (L)-236 (L) and 347-363. [34]

Gerstein, I. S. and R. Jackiw (1969). Anomalies in Ward Identities for Three-Point Functions. *Phys. Rev.* **181**, 1955-1963. [41]

Gilman, F. J. (1969). Sign of the $\pi^0 \to \gamma\gamma$ Decay Amplitude. *Phys. Rev.* **184**, 1964-1965. [35]

Gross, D. and R. Jackiw (1972). Effect of Anomalies on Quasi-Renormalizable Theories. *Phys. Rev. D* **6**, 477-493. [44]

Hagen, C. R. (1969). Derivation of Adler's Divergence Condition from the Field Equations. *Phys. Rev.* **177**, 2622-2623. [39]

Han, M. Y. and Y. Nambu (1965). Three-Triplet Model with Double $SU(3)$ Symmetry. *Phys. Rev.* **139**, B1006-B1010. [35]

Hara, Y. (1964). Unitary Triplets and the Eightfold Way. *Phys. Rev.* **134**, B701-B704. [35]

Iwasaki, Y. (1977). Coupling of the Trace of the Energy-Momentum Tensor to Two Photons. *Phys. Rev. D* **15**, 1172. [47,48]

Jackiw, R. (1970). Field Theoretic Investigations in Current Algebra, in *Lectures on Current Algebra and Its Applications*, S. B. Treiman, R. Jackiw, and D. Gross, eds. (Princeton University Press, Princeton, 1972), pp. 97-254. This was extended to the two articles, R. Jackiw, Field Theoretic Investigations in Current Algebra, and R. Jackiw, Topological

Investigations in Quantum Gauge Theories, in *Current Algebra and Anomalies*, S. Treiman, R. Jackiw, B. Zumino, and E. Witten, eds. (Princeton University Press, Princeton and World Scientific, Singapore, 1985), pp. 81-210 and pp. 211-359. [44]

Jackiw, R. and K. Johnson (1969). Anomalies of the Axial-Vector Current. *Phys. Rev.* **182**, 1459-1469. [36-40]

Jackiw, R. and G. Preparata (1969a). Probes for the Constituents of the Electromagnetic Current and Anomalous Commutators. *Phys. Rev. Lett.* **22**, 975-977. [45]

Jackiw, R. and G. Preparata (1969b). High-Energy Inelastic Scattering of Electrons in Perturbation Theory. *Phys. Rev.* **185**, 1748-1753. [45]

Jackiw, R. and G. Preparata (1969c). T Products at High Energy and Commutators. *Phys. Rev.* **185**, 1929-1940. [45]

Jauch, J. M. and Rohrlich, F. (1955). *The Theory of Photons and Electrons* (Addison-Wesley, Cambridge, MA), Appendix A5-2, pp. 457-461. [31]

Johnson, K. and F. E. Low (1966). Current Algebras in a Simple Model. *Progr. Theor. Phys.* Suppl. **37-38**, 74-93. [36,44,45]

Mack, G. (1968). Partially Conserved Dilatation Current. *Nucl. Phys. B* **5**, 499-507. [47]

Maki, A. (1964). The "fourth" Baryon, Sakata Model and Modified B-L Symmetry. I *Progr. Theoret. Phys.* **31**, 331-332. [35]

Minkowski, P. (1976). On the Anomalous Divergence of the Dilatation Current in Gauge Theories. Univ. Bern preprint, Sept. 1976, archived as KEK scanned version. [48]

Miyamoto, Y. (1965). Three Kinds of Triplet Model, in *Extra Number Supplement of Progress of Theoretical Physics: Thirtieth Anniversary of the Yukawa Meson Theory*, pp. 187-192. [35]

Nambu, Y. (1965) A Systematics of Hadrons in Subnuclear Physics, in *Preludes in Theoretical Physics*, A. de-Shalit, H. Feshbach, and L. Van Hove, eds. (North-Holland, Amsterdam and John Wiley, New York, 1966), pp. 133-142. [35]

Nielsen, N. K. (1977). The Energy-Momentum Tensor in a Non-Abelian Quark Gluon Theory. *Nucl. Phys. B* **120**, 212-220. [48]

Okubo, S. (1969). Sign and Model Dependence of $\pi^0 \to 2\gamma$ Matrix Element. *Phys. Rev.* **179**, 1629-1631. [35]

Pagels, H. and E. Tomboulis (1978). Vacuum of the Quantum Yang-Mills Theory and Magnetostatics. *Nucl. Phys. B* **143**, 485-502. [48,87]

Polkinghorne, J. C. (1958a). Renormalization of Axial Vector Coupling. *Nuovo Cimento* **8**, 179-180. [38]

Polkinghorne, J. C. (1958b). Renormalization of Axial Vector Coupling - II. *Nuovo Cimento* **8**, 781. [38]

Rosenberg, L. (1963). Electromagnetic Interactions of Neutrinos. *Phys. Rev.* **129**, 2786-2788. For an analytic evaluation of the Feynman integrals, see N. N. Achasov, Once More About the Axial Anomaly Pole, *Phys. Lett. B* **287**, 213-215 (1992). [31]

Schwinger, J. (1951). On Gauge Invariance and Vacuum Polarization. *Phys. Rev.* **82**, 664-679. [34,39-41,56]

Schwinger, J. ed. (1958). *Selected Papers on Quantum Electrodynamics* (Dover Publications, New York). [4,39]

Sen, S. (1970). University of Maryland Report No. 70-063. [38]

Sirlin, A. (1978). Current Algebra Formulation of Radiative Corrections in Gauge Theories and the Universality of Weak Interactions. *Rev. Mod. Phys.* **50**, 573-605. See Appendix C, p. 600. [46]

Steinberger, J. (1949). On the Use of Subtraction Fields and the Lifetimes of Some Types of Meson Decay. *Phys. Rev.* **76**, 1180-1186. [34]

Sutherland, D. G. (1967). Current Algebra and Some Non-Strong Mesonic Decays. *Nucl. Phys. B* **2**, 433-440. [32,33,37,47]

Symanzik, K. (1970). Small Distance Behavior in Field Theory and Power Counting. *Commun. Math. Phys.* **18**, 227-246. [39,42,45,48,55,58,59]

Tavkhelidze, A. (1965). Higher Symmetries and Composite Models of Elementary Particles, in *High-Energy Physics and Elementary Particles* (International Atomic Energy Agency, Vienna), pp. 753-762. [35]

Terent'ev, M. V. (1971). Possible Connection between the Amplitudes of the Processes $e^+e^- \to 3\pi$, $\gamma\gamma \to 2\pi$, and $\pi^0 \to 2\gamma$. *ZhETF Pis. Red.* **14**, 140-143 (English translation: *Sov. Phys. JETP Lett.* **14**, 94-96). [43]

't Hooft, G. (1976). Computation of the Quantum Effects due to a Four-Dimensional Pseudoparticle. *Phys. Rev. D* **14**, 3432-3450. [32]

Tung, W.-K. (1969). Equal-Time Commutators and Equations of Motion for Current Densities in a Renormalizable Field-Theory Model. *Phys. Rev.* **188**, 2404-2415. [45]

Vainshtein, A. I. and B. L. Ioffe (1967). Test of Bjorken's Asymptotic Formula in Perturbation Theory. *ZhETF Pis. Red.* **6**, 917-19 (English translation: *Sov. Phys. JETP Lett.* **6**, 341-343). [44]

Veltman, M. (1967). Theoretical Aspects of High Energy Neutrino Interactions. *Proc. Roy. Soc. A* **301**, 107-112. [32,33,37,47]

Veneziano, G. and Y. Yankielowicz (1982). An Effective Lagrangian for the Pure $N = 1$ Supersymmetric Yang-Mills Theory. *Phys. Lett. B* **113**, 231-235. [48]

Weinberg, S. (1973). Theory of Weak and Electromagnetic Interactions, in *Fundamental Interactions in Physics and Astrophysics*, G. Iverson et al., eds. (Plenum Press, New York), pp. 157-186. [44]

Wess, J. and B. Zumino (1971). Consequences of Anomalous Ward Identities. *Phys. Lett. B* **37**, 95-97. [43]

Wilson, K. G. (1969). Non-Lagrangian Models of Current Algebra. *Phys. Rev.* **179**, 1499-1512. [45,47]

Witten, E. (1983). Global Aspects of Current Algebra. *Nucl. Phys. B* **223**, 422-432. [43]

Young, B.-L., T. F. Wong, G. Gounaris, and R. W. Brown (1971). Absence of Second-Order Corrections to the Triangle Anomaly in Quantum Electrodynamics. *Phys. Rev. D* **4**, 348-353. [38]

Zumino, B. (1969). Anomalous Properties of the Axial Vector Current, in *Topical Conference on Weak Interactions Proceedings* CERN 69-7, J. S. Bell, ed. (CERN, Geneva), pp. 361-370. [39]

4. Quantum Electrodynamics

Introduction

My interest in a detailed study of quantum electrodynamics (QED) began during my visit to Cambridge, U.K. in the spring of 1968, when I found the anomalous properties of the axial-vector triangle diagram discussed in Chapter 3. This started me thinking more generally about the properties of fermion loop diagrams, and in particular I wondered whether such diagrams in quantum electrodynamics could lead to an eigenvalue condition for the electric charge, possibly giving an explanation of why the charges of different particle species (such as the electron and proton) are the same in magnitude. This speculation ultimately proved to be wrong, and I look back on the investigations that it inspired with mixed feelings, as being somewhat of a *mis*adventure. On the one hand, my work on aspects of quantum electrodynamics led to a number of important papers with useful results, but on the other hand, my preoccupation with this program kept me from jumping into the emerging area of Yang–Mills unification at the point when much of the interesting theoretical work on non-Abelian theories was being done.

My work on QED divided into three distinct phases, described in the following sections. The first part dealt with a calculation of the process of photon splitting in strong magnetic fields, which served as a warm-up for getting into the study of fermion loop diagrams. After this calculation was completed, I turned to an investigation of the renormalization group properties of QED, using as a tool the newly discovered Callan–Symanzik equations. Finally, in an attempt to get a better formalism for calculating the renormalization group β function contribution from closed loop diagrams, I worked out a compactification of massless QED on the 4-sphere, and applied this formalism to a number of theoretical issues. By the end of this phase, it was clear that developments in non-Abelian gauge theories were the future of the field of particle physics and, through grand unification, offered a compelling way to understand charge quantization, which had been the starting motivation for my interest in electrodynamics. So at this point I set my QED work aside and moved on to some of the phenomenological investigations described in Chapter 5.

Strong Magnetic Field Electrodynamics: Photon Splitting and Vacuum Dielectric Constant

The discovery of pulsars with ultra-strong trapped magnetic fields led to a surge of interest in strong field QED processes, that are unobservably small for attainable laboratory magnetic fields. One of the processes of interest is photon splitting in a constant magnetic field, which is described by a closed electron loop Feynman diagram. When conversations at the Institute turned to whether this reaction could be of relevance in the dynamics of pulsar magnetospheres, my interest in getting into a general study of fermion loop processes in QED made it natural for me to get involved. The initial phase of this study led to a paper (Adler, Bahcall, Callan, and Rosenbluth, 1970, R25), that surveyed the basic features of the photon splitting process. Briefly, the lowest order box diagram makes a vanishing contribution, by an argument using Lorentz invariance and gauge invariance, and so the leading contribution comes from the hexagon diagram, with three insertions of the external magnetic field. (Earlier calculations had overlooked this fact, and so led to the wrong dependence on magnetic field strength.) Using the Heisenberg–Euler effective Lagrangian, we calculated the photon splitting absorption coefficients for the various photon polarization states relative to the magnetic field vector, to leading order in the external magnetic field, for photon energies small relative to the electron mass. We also gave the selection rules that result from the fact that the dielectric constant for the vacuum permeated by a strong magnetic field is different for the different photon polarizations (this was where Marshall Rosenbluth's expertise as a plasma physicist entered in), and made numerical estimates. Some of our results were independently obtained around the same time by Bialynicka-Birula and Bialynicki-Birula (1970).

Again with the aim of getting more experience with QED calculations, I decided to embark on an exact calculation of photon splitting, for arbitrary magnetic fields and for arbitrary photon energies below the pair production threshold. This involved a very lengthy calculation using the proper time method, that Schwinger had first used (Schwinger, 1951) to give an elegant rederivation of the Heisenberg–Euler effective Lagrangian. I derived general formulas for both the photon splitting amplitudes, and the refractive indices needed for the selection rules (in the latter case correcting an earlier result of Minguzzi (1956,1958a,1958b)). I wrote a computer program to numerically evaluate the photon splitting absorption rates, and computed sample results, as well as giving a detailed discussion of possible plasma physics corrections to the selection rules. These results were all reported in a comprehensive article (Adler, 1971, R26) on photon splitting and dispersion in a strong magnetic field.

My overall conclusion was that the leading order calculation from the hexagon diagram gives good order of magnitude estimates, as graphed in Fig. 8 of R26, which plots the ratio of the exact photon splitting absorption coefficient to the hexagon

diagram prediction, versus magnetic field, for photon frequencies equal to zero and equal to the electron mass m. This plot, incidentally, gives a check both on my exact analytic calculation and the numerical work, since the ratio approaches unity for small field strengths, where the hexagon dominates. For magnetic fields of order the "critical field" $B_{CR} = m^2/e \sim 4.41 \times 10^9$ Tesla (4.41×10^{13} Gauss), and photon frequencies of order the electron mass m, the photon splitting mean free path is much shorter than characteristic pulsar magnetosphere depths. However, since the absorption coefficients scale as B^6 for small fields, and since the pulsars known in 1971 tended to have fields of up to a few tenths of B_{CR}, the photon splitting process at that time seemed to be not of great astrophysical importance. Stoneham (1979) published an analytic recalculation of photon splitting by a different method (without numerical evaluation), which as we shall see agreed with my calculation. In an Appendix to his paper, he also improved on my estimate of the very small corrections that arise from the box diagram, when finite opening angles resulting from photon dispersion are taken into account, and we exchanged letters on this aspect of his work. However, after Stoneham's paper, interest in photon splitting waned for quite a number of years.

In the mid 1990's, the discovery of "magnetars", pulsars with fields much higher than the critical field, revived interest in photon splitting. Around April, 1995, John Bahcall told me that recent papers by Mentzel, Berg, and Wunner (1994) and Wunner, Sang, and Berg (1995) claimed that the photon splitting absorption coefficients for energetic photons in strong fields were a factor of 10^4 higher than given in my 1971 paper. If true, this would have had important astrophysical ramifications, so I looked back at my own work, and at the papers of the Wunner group. I was struck by the fact that the Wunner group had not checked to see whether their calculation reproduced the known B^6 dependence of photon splitting for weak fields and low energy photons, a consistency test that, as noted above, I had incorporated into my analytic and numerical work. So I strongly suspected that they had made an error, possibly through a lack of gauge invariance, and wrote a letter to this effect to the Wunner group, while John simultaneously wrote to *Astrophysical Journal Letters*, where their second paper was being considered for publication. Neither of these letters had any effect, and the Wunner, Sang, and Berg paper was published in December, 1995. John Bahcall and Bohdan Paczynski then urged me to make my private misgivings known more publicly. In response, I wrote a short IAS Astrophysics Preprint Series article in January, 1996 (Adler, 1996), expanding on my letter to the Wunner group, and concluding "it is important that their calculation and mine be rechecked by a third party, with the aim of understanding where the discrepancy arises and determining who is right." I submitted this note to the *Astrophysical Journal*, which rejected it.

Although this short note was never published, it had the intended effect as a result of its internal circulation within the IAS. Not long afterwards Christian

Schubert, an IAS visitor at the time, came to my office and said that with new "stringy" Feynman rules with which he was expert, he thought he could repeat in a few days the calculation that had taken me a couple of months by the proper time method. I replied that if he could do that, I would deal with the numerical aspects. A week or two later Christian gave me two equivalent formulas for the photon splitting amplitude obtained by his methods; in the meantime, the Russian group of Baier, Milstein, and Shaisultanov (1996) had produced yet another calculation, which agreed numerically with my 1971 paper. During a short visit to the Institute for Theoretical Physics in Santa Barbara, I wrote programs to directly compare Schubert's two expressions, my 1971 result, Stoneham's 1979 formula, and the analytic formula of the Russian group, all as applied to the allowed polarization case. (The reason for doing this numerically is that an analytic conversion between inequivalent Feynman parameterizations is very difficult, because zero can be written as a multidimensional integral in complicated ways.) The programs showed that the five calculations gave precisely identical amplitudes. This was reported in the paper that I drafted with Schubert on my return to the IAS (Adler and Schubert, 1996, R27). We also posted my computer programs on my web site, and advertised this posting in the paper, so that the community at large could verify what we had done. About a month later, I received an email from Wunner retracting the earlier numerical results of his group, which turned out to result from a single sign error in their computer programs. When this sign error was corrected, the analytic results of Mentzel, Berg and Wunner gave answers that agreed with everyone else, as discussed in Wilke and Wunner (1997). Thus the photon splitting controversy was finally resolved. Subsequently, John Bahcall had me assemble a file of all the relevant papers and correspondence for a post-mortem meeting that he held with the editors of the *Astrophysical Journal*, to analyze and improve the process that that had allowed an incorrect paper to get into print, despite several advance warnings that the results were suspect.

The "Finite QED" Program via the Callan–Symanzik Equations

My comprehensive article on photon splitting was finished in early 1971, and the following summer I returned to my long-standing interest in a study of unresolved issues in the theory of quantum electrodynamics. Johnson, Baker, and Willey (1964), Johnson, Willey, and Baker (1967), and Baker and Johnson (1969, 1971a,b) had written an important series of papers (referred to below as JBW) in which they argued that if QED has a Gell-Mann–Low eigenvalue, then the asymptotic behavior of both the electron and photon propagators would drastically simplify, with the mass term in the electron propagator having power law scaling behavior, and the asymptotic photon propagator behaving, after charge renormalization, as if it had no photon self-energy part. Bill Bardeen and I were both in Aspen for part of the

summer of 1971, and we embarked on a study of QED using the then very new Callan (1970)–Symanzik (1970) equations. Rather than addressing the issue of a possible eigenvalue in QED, we studied the simplified model suggested by the presence of such an eigenvalue, in which the photon propagator is taken as a free propagator with no self-energy part. In this case the β function term, which has a coupling constant derivative, is not present in the Callan–Symanzik equations, and these equations then can be explicitly integrated to give the simple form for the electron propagator found by JBW. These results were described in the paper Adler and Bardeen (1971), R28. In addition to giving results of interest for QED, this paper was one of the first applications of the Callan–Symanzik equations, and was also a motivation for my remarks at the Princeton conference later in 1971 (see R23), in which I suggested a possible connection between an eigenvalue condition in the strong interactions and Bjorken scaling.

After finishing the paper with Bardeen, I turned to a detailed study of the full theory of QED, with photon self-energy parts retained, on which I wrote a comprehensive paper Adler (1972a), R29. This paper had a number of new results. I began with a review of the original Gell-Mann–Low formulation of the renormalization group in QED, and then redid their analysis in terms of the more modern Callan–Symanzik approach, ending up in Eq. (53) with the explicit map between the Callan–Symanzik $\beta(\alpha)$ function and the functions $\psi(\alpha)$ and $q(\alpha)$ that enter into the Gell-Mann–Low formulation. (An implicit form of this map had appeared in Sec. II.3 of Symanzik (1970).) After reviewing the JBW program and the results obtained with Bardeen in R28, I showed by an argument based on the Federbush–Johnson (1960) theorem that *if* there is an eigenvalue in QED, then in the massless limit all $2n$-point current correlation functions must vanish at the eigenvalue. I then went on to show, in an argument that benefited from a conversation with Roger Dashen, that the vanishing of higher correlation functions also implied the vanishing of all coupling constant derivatives of the photon proper self-energy part at the eigenvalue; hence the eigenvalue, if it existed, must be an *infinite order* zero of the one-loop β function. These were all correct results that give the paper an enduring value.

I concluded the paper by proposing that in addition to the standard renormalization group result, in which the eigenvalue plays the role (through running of the coupling) of the unrenormalized fine structure constant α_0, there could be an additional solution, resulting from a fermion-loopwise summation of the theory, in which the eigenvalue plays the role of the physical coupling α. A motivation for this proposal was that the formal power series argument, which shows the equivalence of loopwise summation to the usual renormalization group analysis, could break down in the presence of an essential singularity in the coupling. I then went on to conjecture that loopwise summation with an eigenvalue for α was the mechanism fixing the physical fine structure constant in a uniform manner for all fermion species.

As I have noted in the Introduction to this Chapter, this conjecture turned out to be wrong, and in retrospect my excessive emphasis on it in writing R29 distorted the presentation of an otherwise good paper. At the time key people working on the renormalization group, in particular Gell-Mann, Low, and Wilson, were all very skeptical. Wilson, in particular, remarked at a Princeton seminar that my demonstration of an infinite order zero showed there could be no eigenvalue in QED, and although I was privately annoyed at the time, it is now clear that this was the correct conclusion.

Finally, in an Appendix to my paper, I returned to the electron propagator analysis carried out in R28, this time in a general covariant gauge. This investigation was later reanalyzed in more detail, and improved, in a comprehensive study by Lautrup (1976).

The final paper in this section, Adler, Callan, Gross, and Jackiw (1972), R30, studied the combined implications of the BJL limit, the nonrenormalization of anomalies, and the possible presence of an eigenvalue in QED. This paper, which was initially drafted by Roman Jackiw, grew out of discussions among the authors at Princeton and at the National Accelerator Laboratory. It shows that the following three phenomena are, when taken in combination, incompatible: (1) nonrenormalization of the axial-vector anomaly, (2) the existence of an eigenvalue in QED, (3) validity of naive scale invariant short-distance expansions involving the axial-vector current at the eigenvalue. Since the finite QED program was intended to eliminate the pathologies of QED, through presence of an eigenvalue, this showed that its aims could not be attained, and again cast strong doubt on the existence of an eigenvalue in QED. For later work coming to the same conclusion, and references to more recent literature, see Baker and Johnson (1979), and Acharya and Narayana Swamy (1997). On rereading R30 now, it occurs to me that the argument establishing a relation between the axial-vector anomaly and the Schwinger term given in Section III may be extendable to show that the vanishing of anomalies in axial-vector loop diagrams coupling to four or more photons in QED implies, through similar use of a BJL limit, that the Schwinger term in the two-point function is a c-number. As noted in Chapter 3, this is a result that I was unable to prove, before the advent of the theory of anomalies, in 1966. For another approach to constraining the structure of the Schwinger term, see Jackiw, Van Royen, and West (1970).

Compactification of Massless QED and Applications

The fact that the eigenvalue condition for QED can be studied in the conformally invariant, massless electron theory, led me to study remappings of the Feynman rules for QED that make use of conformal invariance. In Adler (1972b), R31, I showed that the equations of motion and Feynman rules for massless Euclidean QED can be written in terms of equivalent equations of motion and Feynman rules expressed

in terms of coordinates that are confined to the surface of a unit hypersphere in 5-dimensional space (a four-sphere in mathematical terms). For example, letting η^a be the coordinate on the sphere (where a runs from 1 to 5, and $(\eta^a)^2 = 1$), the usual four-vector potential is replaced by a five-vector A^a obeying the constraint (with repeated indices summed) $\eta_a A^a = 0$, and the electromagnetic field strength is replaced by a three-index tensor $F_{abc} = L_{ab}A_c + L_{bc}A_a + L_{ca}A_b$, with L_{ab} the 5-space rotation generators. This tensor has a two-index dual $\hat{F}_{ab}$, and the Maxwell equations become $L_{ab}F_{abc} = 2eJ_c$, $L_{ab}\hat{F}_{bc} = \hat{F}_{ac}$. The corresponding $O(5)$ covariant Feynman rules are given in Table I of R31. The result of this transformation of the theory is an explicit demonstration that massless QED can be compactified, so that there are only ultraviolet divergences (corresponding to points approaching each other on the surface of the sphere, where it becomes tangent to Euclidean 4-space), but no infrared divergences. The $O(5)$ rules, however, are not manifestly conformal invariant; in a later section of the paper I showed that they are related, by a projective transformation, to a manifestly conformal invariant (but non-compact) $O(5,1)$ formalism that was introduced earlier by Dirac (1936).

In two subsequent papers I further developed and applied the $O(5)$ covariant formalism. In Adler (1973) I showed that the usual Feynman path integral takes the form of an amplitude integral, constructed as an infinite product of individually well-defined ordinary integrals over coefficients appearing in the hyperspherical harmonic expansion of the electromagnetic potential A_a. In the paper Adler (1974), R32, I used the amplitude integral formalism to study a simple model, in which only a single photon mode of the form $A_a \propto v_{1a}\eta \cdot v_2 - v_{2a}\eta \cdot v_1$, with $v_{1,2}$ orthogonal unit vectors, is retained. The external field Fredholm determinant or vacuum persistence amplitude $\Delta(eA) = \det(i\gamma \cdot \partial + e\gamma \cdot A)$ could then be studied by exploiting the $O(3) \times O(2)$ residual symmetry of this model, which permits the external field problem to be reduced to a set of two coupled first order ordinary differential equations, with a Wronskian equal to the Fredholm determinant. A significant result coming out of the analysis of this model was that the renormalized Fredholm determinant is an entire function of order four as eA becomes infinite in a general complex direction. This played a role in a subsequent discussion of asymptotic estimates in perturbative QED, as discussed in the paper of Balian, Itzykson, Zuber, and Parisi (1978), which followed up on an earlier paper of Itzykson, Parisi, and Zuber (1977). Whereas extrapolation from the solvable case of a constant field strength $F_{\mu\nu}$ suggested that the order of the Fredholm determinant is two, my solvable example showed that two cannot be the correct answer for general vector potentials. Balian et al. noted this and then went on to present further arguments for the determinant being of order four in four-dimensional spacetime, or more generally of order D in D-dimensional spacetime. This in turn had important implications for their study of asymptotic behavior of the perturbation series in QED. The subject of the order of the Fredholm determinant was further developed by Bogomolny. In an initial

paper by Bogomolny and Fateyev (1978), the case of fields with an $O(3) \times O(2)$ symmetry group that I had initiated in R32 was taken up again, and an asymptotic formula for the Fredholm determinant was obtained. In a subsequent paper, Bogomolny (1979) showed that this asymptotic formula, and a similar formula obtained by Balian et al. for another special case, could be extended to the general result $\lim_{e\to i\infty} \Delta(eA) = (e^4/12\pi^2) \int d^4x \big((A_\mu)^2\big)^2$, provided A_μ is chosen to obey the non-linear gauge condition $\partial_\mu(A_\mu A^2) = 0$. Thus the order four result that I found in my "one-mode" model in fact gave the correct general answer for QED.

A further application of the $O(5)$ formalism for QED emerged after the discovery of the instanton solution to the Yang–Mills field equations. Jackiw and Rebbi (1976) showed that the one-instanton solution is invariant under an $O(5)$ subgroup of the full conformal group, and hence can be rewritten in an elegant way in terms of the $O(5)$ formulation of electrodynamics, as extended to non-Abelian gauge fields. Letting α_a be the $O(5)$ equivalent of the Dirac γ matrices, and $\gamma_{ab} = (i/4)[\alpha_a, \alpha_b]$, a matrix-valued vector potential A_a obeying the constraint $\eta \cdot A = 0$ can be immediately constructed as $A_a = C\eta_b\gamma_{ab}$. Jackiw and Rebbi showed that when this vector potential is substituted into the Yang–Mills field equation as expressed in the non-Abelian extension of the $O(5)$ formalism, one gets a cubic equation for the coefficient C, two roots of which give pure gauge potentials with vanishing field strengths, but the third root of which gives the instanton! Thus, I had missed a significant opportunity in not pursuing the question, raised at least once when I gave seminars, of what the non-Abelian generalization of the $O(5)$ formalism was like. A variant of the non-Abelian $O(5)$ formalism was subsequently applied by Belavin and Polyakov (1977), with corrections by Ore (1977), to give a recalculation of the Fredholm determinant in an instanton background that was first computed by 't Hooft (1976).

References for Chapter 4

Acharya, R. and P. Narayana Swamy (1997). No Eigenvalue in Finite Quantum Electrodynamics. *Int. J. Mod. Phys. A* **12**, 3799-3809. [60]

Adler, S. L. (1971) R26. Photon Splitting and Photon Dispersion in a Strong Magnetic Field. *Ann. Phys.* **67**, 599-647. Pages 599-601, 609-613, 621-622, and 634-644 are reprinted here. [56-58]

Adler, S. L. (1972a) R29. Short-Distance Behavior of Quantum Electrodynamics and an Eigenvalue Condition for α. *Phys. Rev. D* **5**, 3021-3047. [59,60]

Adler, S. L. (1972b) R31. Massless, Euclidean Quantum Electrodynamics on the 5-Dimensional Unit Hypersphere. *Phys. Rev. D* **6**, 3445-3461. [60-62]

Adler, S. L. (1973). Massless Electrodynamics on the 5-Dimensional Unit Hypersphere: An

Amplitude-Integral Formulation. *Phys. Rev. D* **8**, 2400-2418. [61]

Adler, S. L. (1974) R32. Massless Electrodynamics in the One-Photon-Mode Approximation. *Phys. Rev. D* **10**, 2399-2421. [61,62]

Adler, S. L. (1996). Comment on "Photon Splitting in Strongly Magnetized Objects Revisited". IASSNS-AST 96/4 (unpublished). [57]

Adler, S. L., J. N. Bahcall, C. G. Callan, and M. N. Rosenbluth (1970) R25. Photon Splitting in a Strong Magnetic Field. *Phys. Rev. Lett.* **25**, 1061-1065. [56]

Adler, S. L. and W. A. Bardeen (1971) R28. Quantum Electrodynamics without Photon Self-Energy Parts: An Application of the Callan-Symanzik Scaling Equations. *Phys. Rev. D* **4**, 3045-3054. [58-60]

Adler, S. L., C. G. Callan, D. J. Gross, and R. Jackiw (1972) R30. Constraints on Anomalies. *Phys. Rev. D* **6**, 2982-2988. [60]

Adler, S. L. and C. Schubert (1996) R27. Photon Splitting in a Strong Magnetic Field: Recalculation and Comparison with Previous Calculations. *Phys. Rev. Lett.* **77**, 1695-1698. [58]

Baier, V. N., A. I. Milstein, and R. Zh. Shaisultanov (1996). Photon Splitting in a Very Strong Magnetic Field. *Phys. Rev. Lett.* **77**, 1691-1694. [58]

Baker, M. and K. Johnson (1969). Quantum Electrodynamics at Small Distances. *Phys. Rev.* **183**, 1292-1299. [58]

Baker, M. and K. Johnson (1971a). Asymptotic Form of the Electron Propagator and the Self-Mass of the Electron. *Phys. Rev. D* **3**, 2516-2526. [58]

Baker, M. and K. Johnson (1971b). Simplified Equation for the Bare Charge in Renormalized Quantum Electrodynamics. *Phys. Rev. D* **3**, 2541-2542. [58]

Baker, M. and K. Johnson (1979). Applications of Conformal Symmetry in Quantum Electrodynamics. *Physica A* **96**, 120-130. [60]

Balian, R., C. Itzykson, J. B. Zuber, and G. Parisi (1978). Asymptotic Estimates in Quantum Electrodynamics. II. *Phys. Rev. D* **17**, 1041-1052. [61,62]

Belavin, A. A. and A. M. Polyakov (1977). Quantum Fluctuations of Pseudoparticles. *Nucl. Phys. B* **123**, 429-444. [62]

Bialynicka-Birula, Z. and I. Bialynicki-Birula (1970). Nonlinear Effects in Quantum Electrodynamics. Photon Propagation and Photon Splitting in an External Field. *Phys. Rev. D* **2**, 2341-2345. [56]

Bogomolny, E. B. (1979). Large-Complex-Charge Behavior of the Dirac Determinant. *Phys. Lett. B* **86**, 199-202. [62]

Bogomolny, E. B. and V. A. Fateyev (1978). The Dyson Instability and Asymptotics of the Perturbation Series in QED. *Phys. Lett. B* **76**, 210-212. [62]

Callan, C. G. (1970). Broken Scale Invariance in Scalar Field Theory. *Phys. Rev. D* **2**, 1541-1547. [39,42,45,48,55,58,59]

Dirac, P. A. M. (1936). Wave Equations in Conformal Space. *Ann. Math.* **37**, 429-442. [61]

Federbush, P. G. and Johnson, K. A. (1960). Uniqueness Property of the Twofold Vacuum Expectation Value. *Phys. Rev.* **120**, 1926. [59]

Itzykson, C., G. Parisi, and J-B. Zuber (1977). Asymptotic Estimates in Quantum Electrodynamics. *Phys. Rev. D* **16**, 996-1013. [61]

Jackiw, R. and C. Rebbi (1976). Conformal Properties of a Yang-Mills Pseudoparticle. *Phys. Rev. D* **14**, 517-523. [62]

Jackiw, R., R. Van Royen, and G. B. West (1970). Measuring Light-Cone Singularities. *Phys. Rev. D* **2**, 2473-2485. See especially Sec. II B. [60]

Johnson, K., M. Baker, and R. Willey (1964). Self-Energy of the Electron. *Phys. Rev.* **136**, B1111-B1119. [58]

Johnson, K., R. Willey, and M. Baker (1967). Vacuum Polarization in Quantum Electrodynamics. *Phys. Rev.* **163**, 1699-1715. [58]

Lautrup, B. (1976). Renormalization Constants and Asymptotic Behavior in Quantum Electrodynamics. *Nucl. Phys. B* **105**, 23-44. [60]

Mentzel, M., D. Berg, and G. Wunner (1994). Photon Splitting in Strong Magnetic Fields. *Phys. Rev. D* **50**, 1125-1139. [57,58]

Minguzzi, A. (1956). Non-Linear Effects in the Vacuum Polarization. *Nuovo Cimento* **4**, 476-486. [56]

Minguzzi, A. (1958a). Non Linear Effects in the Vacuum Polarization (II). *Nuovo Cimento* **6**, 501-511. [56]

Minguzzi, A. (1958b). Causality and Vacuum Polarization due to a Constant and a Radiation Field. *Nuovo Cimento* **9**, 145-153. [56]

Ore, F. R. (1977). How to Compute Determinants Compactly. *Phys. Rev. D* **16**, 2577-2580. [62]

Schwinger, J. (1951). On Gauge Invariance and Vacuum Polarization. *Phys. Rev.* **82**, 664-679. [34,39-41,56]

Stoneham, R. J. (1979). Photon Splitting in the Magnetized Vacuum. *J. Phys. A: Math. Gen.* **12**, 2187-2203. [57,58]

Symanzik, K. (1970). Small Distance Behavior in Field Theory and Power Counting. *Commun. Math. Phys.* **18**, 227-246. [39,42,45,48,55,58,59]

't Hooft, G. (1976). Computation of the Quantum Effects due to a Four-Dimensional Pseudoparticle. *Phys. Rev. D* **14**, 3432-3450. [62]

Wilke, C. and Wunner, G. (1997). Photon Splitting in Strong Magnetic Fields: Asymptotic Approximation Formulas versus Accurate Numerical Results. *Phys. Rev. D* **55**, 997-1000. [58]

Wunner, G., R. Sang, and D. Berg (1995). Photon Splitting in Strongly Magnetized Cosmic Objects – Revisited. *Astrophys. J.* **455**, L51-L53. [57]

5. Particle Phenomenology and Neutral Currents

Introduction

Much of the work described in Chapters 2 and 3 on soft pion theorems, sum rules, anomalies, and neutrino reactions falls in the category of phenomenology, but both the interrelations between different aspects of this research, and the chronology, suggested that it be discussed earlier. Even before this work was done, I wrote my first particle phenomenology paper in collaboration with my first year Princeton graduate school roommate, and former Harvard classmate, Alfred Goldhaber (Adler and Goldhaber, 1963). In this paper we analyzed the possibility of using the deuteron to provide a polarized proton target, by determining the polarization of the recoiling spectator neutron through its scattering on He^4. Although perhaps feasible, this proposal was never implemented, and much better methods for directly obtaining polarized targets are now available. After I completed the work on quantum electrodynamics described in Chapter 4, I returned to phenomenology in a number of papers written, or conceived, during visits to the National Accelerator Laboratory (subsequently renamed the Fermi National Accelerator Laboratory, or Fermilab), and continued with related work in a number of papers written at the IAS. I discuss the earlier work done at Fermilab in the first section that follows, and then in the second section take up work at both Fermilab and the IAS relating to neutral currents.

Visits to Fermilab

When the National Accelerator Laboratory was inaugurated, my former thesis advisor Sam Treiman was brought in, on a succession of leaves from Princeton starting in 1970, to serve as temporary head of the Theory Group, with the charge of setting it up and recruiting a permanent head. Subsequently, Ben Lee was hired to be the permanent head of the Theory Group. During this period many theorists from outside institutions were invited to be term time and/or summer visitors, and as part of this program I made a series of visits to Fermilab, and wrote a number of phenomenological papers growing out of discussions with people there.

As already noted in Chapter 3, during a 1971 visit to Fermilab I collaborated with Lee, Treiman, and Tony Zee to study the anomaly-based prediction for the process $\gamma\gamma \to 3\pi$, described in the paper R20. This was applied in a subsequent paper

that I wrote with Glennys Farrar and Treiman (Adler, Farrar, and Treiman, 1972, R33) to an analysis of the contribution of three pion intermediate states to the rare kaon decay $K_L \to \mu^+\mu^-$. The background for this study was what was then called the "$K_L \to \mu^+\mu^-$ puzzle", the fact that experiment had not detected this kaon decay mode at a level considerably below that given by a unitarity bound based on the assumed dominance of a two photon intermediate state in the absorptive part of the decay amplitude. There were thus two possibilities, either an experimental problem, or destructive interference with another intermediate state, for which the three pion intermediate state was a prime candidate. Aviv and Sawyer (1971) had proposed to use soft pion methods to estimate the three pion contribution, and had concluded that the contribution was much too small to be relevant. However, the Aviv–Sawyer analysis used an expression for the $3\pi \to \gamma\gamma$ amplitude which had been shown in R20 to be incorrect. In R33, we estimated the three pion contribution by using the corrected $3\pi \to \gamma\gamma$ amplitude calculated in R20, but still found that it gave much too small a contribution to explain the lack of observed $K_L \to \mu^+\mu^-$ events. Similar conclusions, again using the results of R20, were reached independently by Pratap, Smith, and Uy (1972). Ultimately, the origin of the "$K_L \to \mu^+\mu^-$ puzzle" turned out to be experimental, and this decay mode has now been seen in a number of experiments, with the Particle Data Group giving an average value for $\Gamma(\mu^+\mu^-)/\Gamma_{\mathrm{TOT}}$ of $\sim 7.2 \times 10^{-9}$, as compared with the theoretical unitarity lower bound of 7.0×10^{-9} based on the current $K_L \to \gamma\gamma$ branching ratio.

During the years 1973-1974, my Fermilab visits led to papers in two separate areas, searches for neutral currents in weak pion production, and the analysis of what was then a discrepancy between theory and experiment in μ-mesic atom x-ray spectra. I will take up this second area first, because the neutral current work leads directly into the papers discussed in the next section. My interest in the μ-mesic atom discrepancy was stimulated by my earlier work on quantum electrodynamics, since an eigenvalue in QED could show up as deviations from the standard perturbation theory predictions for vacuum polarization effects. Thinking about tests for vacuum polarization discrepancies in QED led me to think more generally about other aspects of vacuum polarization, in particular the predictions for the ratio $R(s) = \sigma(e^+e^- \to \text{hadrons}; s)/\sigma(e^+e^- \to \mu^+\mu^-; s)$ in various models for quark structure of hadrons. This offshoot of the QED work led to results that are still used today, introduced in the paper Adler (1974a), R34, dealing with "Some simple vacuum-polarization phenomenology...". My basic observation was that whereas R is measured in the timelike region, the natural place to compare experiment with scaling predictions of various theories is in the spacelike region, where (since there are no threshold effects) one might expect an early or "precocious" onset of scaling. Rather than directly using the dispersion relation for the vacuum polarization part to calculate the spacelike continuation, I proposed using its first derivative, and so

defined a function

$$T(-s) = \int_{4m_\pi^2}^{\infty} \frac{duR(u)}{(s+u)^2} \quad .$$

This function is the one for which parton models and QCD most directly make predictions, and since it is positive definite and involves a strongly convergent integral (for R approaching a constant), the experimentally inaccessible high energy tail has a known sign and a magnitude that can be bounded. For a parton model in which R asymptotically approaches a constant C, one has $T(-s) \sim C/s$ as $s \to \infty$, and a similar formula holds in QCD with a known logarithmic correction. The paper R34 used the function $T(-s)$ to propose a test of the colored quark hypothesis. Subsequently, De Rújula and Georgi (1976) used a modified version of this idea, defining $D(s) = sT(-s)$, to analyze the new SPEAR data. They found that the original colored quark model was excluded, and among various viable possibilities, noted that "the standard model with charm is acceptable if heavy leptons are produced," a conclusion that was borne out by experiment with the subsequent discovery of the τ lepton. Shortly afterwards, Poggio, Quinn, and Weinberg (1976) proposed a generalized method in which the derivative of the hadronic vacuum polarization that I had used is replaced by a finite difference between the hadronic vacuum polarization values at points a distance $\pm i\Delta$ from the timelike real axis, leading to a "smeared" average of $R(s)$ that retains sensitivity to threshold effects. Recently, my original method, generally in the form $D(s)$ used by De Rújula and Georgi, has been revived under the name of the "Adler function", in a number of papers; see, for example, Broadhurst and Kataev (1993), Kataev (1996); Peris, Perrottet, and de Rafael (1998); Beneke (1999); Eidelman, Jegerlehner, Kataev, and Veretin (1999); Kataev (1999); Cvetič, Lee, and Schmidt (2001); Cvetič, Dib, Lee, and Schmidt (2001); Milton, Solovtsov, and Solovtsova (2001); and Dorokhov (2004).

In the second part of R34 I examined what was then a discrepancy between theory and experiment in μ-mesic atom x-ray transition energies, under the assumption that (if real) the discrepancy arose from a nonperturbative correction $\delta\rho$ to the vacuum polarization absorptive part. Assuming that $\delta\rho$ is positive, or positive and monotonic, I derived lower bounds on the corresponding deviation that would be expected in $a_\mu = \frac{1}{2}(g_\mu - 2)$. For instance, if $\delta\rho$ is assumed positive and monotonic, comparison of the kernels that weight ρ in the formulas for the x-ray transition energies and for a_μ gives the bound $\delta a_\mu \leq -(0.98 \pm 0.18) \times 10^{-7}$. In a follow-up paper with Roger Dashen and Sam Treiman (Adler, Dashen, and Treiman, 1974) we discussed other tests for a nonperturbative vacuum polarization contribution, and also placed bounds on the mass of a light scalar meson that could be invoked to explain the x-ray discrepancy. A few months later, Barbieri (1975) extended the method of R34 to show that precision measurements of the $(\mu\,^4\text{He})^+$ system were already at variance, within the vacuum polarization deviation or scalar meson

exchange hypotheses, with the supposed x-ray discrepancy. A later paper of Barbieri and Ericson (1975) gave additional evidence against the scalar meson explanation for the x-ray discrepancy. In the meantime, during 1975 and the few years following, there were a number of experimental developments, reviewed in detail in Borie and Rinker (1982), as a result of which the muonic x-ray discrepancy was eliminated. Incidentally, the current theoretical and experimental values of a_μ differ by a few parts in 10^{-9}, well below the lower bounds on δa_μ inferred in R34 from the μ-mesic atom x-ray data at that time, giving an additional indication that that the purported x-ray discrepancy was an experimental artifact.

Neutral Currents

The existence of weak neutral currents is a principal prediction of the Glashow–Weinberg–Salam electroweak theory, and commanded much attention in the 1970s. Failure to find weak neutral currents would have falsified the electroweak theory, and on the other hand, detection of weak neutral currents would give a value for the electroweak mixing angle θ_W, which in turn determines the masses of the heavy intermediate bosons of the theory. As a result of my thesis work on weak pion production, it was natural for me to get interested in theoretical estimates of the neutral current weak pion production channels $\nu + N_i \to \nu + \pi + N_f$, with $N_{i,f}$ a nucleon (either a neutron or proton) and with π a pion of appropriate charge. In July 1972, a collaboration with Wonyong Lee as spokesman proposed a study of weak neutral currents in both the purely leptonic and the pion production channels at the Brookhaven AGS accelerator, and a copy of their proposal is in my files. Through this, and through related correspondence of Ben Lee with Sam Treiman, I got interested in doing detailed calculations for this process, and over the next few years was in frequent touch with the experimental group for which Wonyong Lee was spokesman.

My initial papers were motivated by the fact that preliminary estimates of neutral current weak pion production by Ben Lee (1972) appeared to conflict with experiments in complex nuclei reported by Wonyong Lee (1972), subject to two caveats. The first caveat was that Ben Lee's static model estimates didn't include $I = 1/2$ contributions to weak pion production, and the second caveat was that nuclear charge exchange corrections could be important, as noted by Perkins (1972). The first of these issues was dealt with in a short paper Adler (1974b), R35, where I used my model of R15, as adapted to the neutral current case, to estimate the effects of including the nonresonant isospin 1/2 channels, and concluded that they had little effect on Ben Lee's estimate from the dominant isospin 3/2 channel. The second issue was dealt with in a paper on nuclear charge exchange corrections to pion production in the $\Delta(1232)$ region, that I wrote in collaboration with Shmuel Nussinov and Emmanuel Paschos (Adler, Nussinov, and Paschos, 1974, R36). In this paper,

we estimated the effects of multiple charge exchange scattering on pion production in nuclear targets, using an extension of techniques used by Fermi and others to calculate multiple neutron scattering in the early days of neutron physics. A considerable part of the fun of writing this paper was learning about this older work on neutron physics, and feeling a sense of continuity between current concerns of weak interaction physics and the quite differently motivated work of an earlier generation. In addition to giving analytic formulas, we tabulated various results for the case of a $_{13}Al^{27}$ target, as appropriate to experiments with aluminum spark chamber plates. In R36, we made the simplifying assumption of an isotopically neutral target (that is, equal numbers of neutrons and protons), which is exact for $_{6}C^{12}$, and a good approximation for aluminum. In a follow up paper (Adler, 1974c), I extended the model to nuclear targets with a neutron excess. As can be seen from Table II of R36, charge exchange corrections are sizable, and in our model typically reduce the ratio of neutral current to charged current π^0 production by about 40%.

My next paper on neutral currents was motivated by the fact that preliminary results of an experiment on weak pion production in hydrogen at Argonne National Laboratory showed a cluster of neutral current events just above threshold. In this kinematic regime soft pion methods should apply, allowing one to relate threshold neutral current weak pion production in the standard electroweak theory to the elastic neutral current cross section for $\nu + p \rightarrow \nu + p$. Using this relation, I showed in Adler (1974d), R37 that one could place bounds on the expected number of neutral current pion production events in the threshold region, with the Argonne results exceeding these bounds. Thus, there seemed to be stronger neutral current weak pion production than suggested by the $SU(2) \times U(1)$ electroweak theory.

Subsequent events then proceeded on several parallel tracks. In a follow-up paper to R37, published as Adler (1975a), R38, I used the full apparatus of my weak pion production calculation of R15 to extend the neutral current calculation above the threshold region to include the regime where $\Delta(1232)$ production dominates. This analysis reinforced the conclusions about the preliminary Argonne data already reached in R37. Simultaneously, with a large group of postdocs at the Institute, I embarked on a study of weak pion production in alternative models of neutral currents with scalar, pseudoscalar, and tensor currents, and also with so-called "second class" (abnormal G-parity) currents. Additionally, in Adler, Karliner, Lieberman, Ng, and Tsao (1976), we did a detailed study of isospin-1/2 resonance production by V, A neutral currents. Perhaps the one part of the group effort on alternative current structures to have lasting value was a calculation of nucleon to nucleon and pion to pion matrix elements of scalar, pseudoscalar, and tensor current densities, using all the theoretical tools then at our disposal: flavor SU_3 and chiral $SU_3 \times SU_3$ symmetries, the quark model, and the MIT "bag" model. The results of these calculations were checked by several of us, and tabulated in Adler, Colglazier, Healy, Karliner, Lieberman, Ng, and Tsao (1975), R39; they were subsequently relevant for

estimates of the coupling to nucleons of hypothetical scalar and pseudoscalar particles, such as axions. The main part of the group effort was a current algebra soft pion production calculation for the alternative current case, which involved extensive algebra and computer work. From this, we found that one could explain roughly half of the reported Argonne threshold events with currents of scalar, pseudoscalar, and tensor type, by allowing some deviations from the matrix element estimates of R39, as I reported at the January, 1975 Coral Gables Conference (Adler, 1975b). In the meantime, the Argonne group reexamined possible background problems affecting their preliminary results, with the result that they ultimately discounted the cluster of pion production events near threshold. So by September of 1975, when I reviewed the subject of gauge theories and neutrino interactions at a conference at Northeastern University (Adler, 1976a), the electroweak theory predictions for neutral current weak pion production, following from purely V and A currents, were no longer in conflict with experiment. This conclusion was reinforced by a subsequent detailed analysis by Monsay (1978) of neutral current weak pion production, using my model together with the charge exchange corrections of R36.

In the summer of 1975 I lectured on neutrino interactions and neutral currents at the Sixth Hawaii Topical Conference on Particle Physics, and gave a comprehensive survey of neutral current phenomenology based on parton model methods, soft pion theorems, and quark model calculations of baryon static properties. This appeared both in the conference proceedings (Adler, 1976b) and again in a tenth year anniversary volume selecting highlights from the preceding summer schools (Pakvasa and Tuan, 1982). My hope in preparing the 1975 lectures was that surveying all available tools would hasten the day when one could determine electroweak parameters based on using all available data for a global fit, instead of doing piecemeal fits channel-by-channel. Such a global fit was carried out a few years later by Abbott and Barnett (1978a,b), who included four types of data: deep inelastic neutrino scattering $\nu N \to \nu X$, elastic neutrino-proton scattering $\nu p \to \nu p$, neutrino induced inclusive pion production $\nu N \to \nu \pi X$, and neutrino induced exclusive pion production $\nu N \to \nu \pi N$. For the exclusive pion process, they employed my weak pion production calculation of R15 as extended to neutral currents in R38, using test data that I ran for them from my programs as benchmarks to help debug their programming. Their results were, in the words of their letter Abstract, "for the first time, a unique determination of the weak neutral-current couplings of u and d quarks. Data for exclusive pion production are a crucial new input in this analysis." Their multi-channel fit gave the first full confirmation that the Glashow–Weinberg–Salam model, with $\sin^2 \theta_W$ between 0.22 and 0.30, was in agreement with the experimental up and down quark neutral current coupling parameters. To me, the Abbott–Barnett analysis was valued recompense for the several years of hard calculation and scholarly attention to detail that I had put into the subject of weak pion production.

References for Chapter 5

Abbott, L. F. and R. M. Barnett (1978a). Determination of the Weak Neutral-Current Couplings. *Phys. Rev. Lett.* **40**, 1303-1306. [70]

Abbott, L. F. and R. M. Barnett (1978b). Quark and Lepton Couplings in the Weak Interactions. *Phys. Rev. D* **18**, 3214-3229. [70]

Adler, S. L. (1974a) R34. Some Simple Vacuum-Polarization Phenomenology: $e^+e^- \to$ Hadrons; The Muonic-Atom X-Ray Discrepancy and $g_\mu - 2$. *Phys. Rev. D* **10**, 3714-3728. [66-68]

Adler, S. L. (1974b) R35. $I = \frac{1}{2}$ Contributions to $\nu_\mu + N \to \nu_\mu + N + \pi^0$ in the Weinberg Weak-Interaction Model. *Phys. Rev. D* **9**, 229-230. [68]

Adler, S. L. (1974c). Pion Charge-Exchange Scattering in the (3,3)-Resonance Region in Nuclei with a Neutron Excess. *Phys. Rev. D* **9**, 2144-2150. [69]

Adler, S. L. (1974d) R37. Application of Current Algebra Techniques to Neutral-Current–Induced Threshold Pion Production. *Phys. Rev. Lett.* **33**, 1511-1514. [69]

Adler, S. L. (1975a) R38. Application of Current-Algebra Techniques to Soft-Pion Production by the Weak Neutral Current: *V*, *A* Case. *Phys. Rev. D* **12**, 2644-2655. [69,70]

Adler, S. L. (1975b). Theoretical Interpretation of Recent Neutral Current Results, in *Theories and Experiments in High-Energy Physics*, Proceedings of Orbis Scientiae, the second Coral Gables Conference at the University of Miami, January 20-24, 1975, B. Kursunoglu, chair, A. Perlmutter and S. M. Widmayer, eds. (Plenum Press, New York), pp. 297-327. [70]

Adler, S. L. (1976a). Gauge Theories and Neutrino Interactions, in *Gauge Theories and Modern Field Theory*, R. Arnowitt and P. Nath, eds. (MIT Press, Cambridge, MA), pp. 127-160. [70]

Adler, S. L. (1976b). Neutrino Interaction Phenomenology and Neutral Currents, in *Proceedings of the Sixth Hawaii Topical Conference in Particle Physics (1975)*, P. N. Dobson, S. Pakvasa, V. Z. Peterson, and S. F. Tuan, eds. (University of Hawaii, Manoa/Honolulu), pp. 1-207. [70]

Adler, S. L., E. W. Colglazier, J. B. Healy, I. Karliner, J. Lieberman, Y. J. Ng, and H.-S. Tsao (1975) R39. Renormalization Constants for Scalar, Pseudoscalar, and Tensor Currents. *Phys. Rev. D* **11**, 3309-3318. [69,70]

Adler, S. L., R. F. Dashen, and S. B. Treiman (1974). Comments on Proposed Explanations for the Muonic-Atom X-Ray Discrepancy. *Phys. Rev. D* **10**, 3728-3735. [67]

Adler, S. L., G. R. Farrar, and S. B. Treiman (1972) R33. Three-Pion States in the $K_L \to \mu^+\mu^-$ Puzzle. *Phys. Rev. D* **5**, 770-772. [66]

Adler, S. L. and A. S. Goldhaber (1963). Use of the Deuteron to Provide a Polarized Proton Target. *Phys. Rev. Lett.* **10**, 448-450. [65]

Adler, S. L., I. Karliner, J. Lieberman, Y. J. Ng, and H.-S. Tsao (1976). Isospin-$\frac{1}{2}$ Nucleon-Resonance Production by a V, A Weak Neutral Current. *Phys. Rev. D* **13**, 1216-1233. [69]

Adler, S. L., S. Nussinov, and E. A. Paschos (1974) R36. Nuclear Charge-Exchange Corrections to Leptonic Pion Production in the (3,3)-Resonance Region. *Phys. Rev. D* **9**, 2125-2143. [68-70]

Aviv, R. and R. F. Sawyer (1971). Three-Pion Intermediate State and the $K_L^0 \to \mu^+\mu^-$ Puzzle. *Phys. Rev. D* **4**, 2740-2742. [66]

Barbieri, R. (1975). Vacuum Polarization Phenomenology: The μ Mesic Atom X-Ray Discrepancy and the $2P_{\frac{3}{2}} - 2S_{\frac{1}{2}}$ Separation in the $(\mu\,^4\mathrm{He})^+$ System. *Phys. Lett. B* **56**, 266-270. [67]

Barbieri, R. and T. E. O. Ericson (1975). Evidence against the Existence of a Low Mass Scalar Boson from Neutron-Nucleus Scattering. *Phys. Lett. B* **57**, 270-272. [68]

Beneke, M. (1999). Renormalons. *Physics Reports* **3**, 1-142, Sec. 2.2. Beneke uses the notation $D(Q^2)$ but does not include an additional factor of Q^2. [67,79]

Borie, E. and G. A. Rinker (1982). The Energy Levels of Muonic Atoms. *Rev. Mod. Phys.* **54**, 67-118. See particularly pp. 105-108. [68]

Broadhurst, D. J. and A. L. Kataev (1993). Connections between Deep-Inelastic and Annihilation Processes at Next-to-Next-to-Leading Order and Beyond. *Phys. Lett. B* **315**, 179-187. [67]

Cvetič, G., C. Dib, T. Lee, and I. Schmidt (2001). Resummation of the Hadronic Tau Decay Width with the Modified Borel Transform Method. *Phys. Rev. D* **64**, 093016. [67]

Cvetič, G., T. Lee, and I. Schmidt (2001). Resummations with Renormalon Effects for the Leading Hadronic Contribution to the Muon $g_\mu - 2$. *Phys. Lett. B* **520**, 222-232. [67]

De Rújula, A. and H. Georgi (1976). Counting Quarks in e^+e^- Annihilation. *Phys. Rev. D* **13**, 1296-1301. [67]

Dorokhov, A. (2004). Adler Function and Hadronic Contribution to the Muon $g - 2$ in a Nonlocal Chiral Quark Model. *Phys. Rev. D* **70**, 094011. [67]

Eidelman, S., F. Jegerlehner, A. L. Kataev, and O. Veretin (1999). Testing Non-Perturbative Strong Interaction Effects via the Adler Function. *Phys. Lett. B* **454** 369-380. [67]

Kataev, A. L. (1996). The Generalized Crewther Relation: The Peculiar Aspects of the Analytical Perturbative QCD Calculations; arXiv: hep-ph/9607426. [67]

Kataev, A. L. (1999). Adler Function from $R^{e^+e^-}(s)$ Measurements: Experiments vs QCD Theory, in *Moscow 1999, Particle Physics at the Start of the New Millennium*, pp. 43-52; arXiv: hep-ph/9906534. [67]

Lee, B. W. (1972). The Process $\nu_\mu + p \to \nu_\mu + p + \pi^0$ in Weinberg's Model of Weak Interactions. *Phys. Lett. B* **40**, 420-422. [68]

Lee, W. (1972) Experimental Limit on the Neutral Current in the Semileptonic Processes. *Phys. Lett. B* **40**, 423-425. [68]

Milton, K. A., I. L. Solovtsov, and O. P. Solovtsova (2001). Adler Function for Light Quarks in Analytic Perturbation Theory. *Phys. Rev. D* **64**, 016005. [67]

Monsay, E. H. (1978). Single-Pion Production by the Weak Neutral Current. *Phys. Rev. D* **18**, 2277-2289. [70]

Pakvasa, S. and S. F. Tuan, eds. (1982). *Selected Lectures, Hawaii Topical Conference in Particle Physics* (World Scientific, Singapore). My 1975 lectures Adler (1976b) are reprinted in Vol. 2 of this compilation, pp. 499-705. [70]

Peris, S., M. Perrottet, and E. de Rafael (1998). Matching Long and Short Distances in Large-N_c QCD. *JHEP* **05**, 011. [67]

Perkins, D. H. (1972) Neutrino Interactions, in *Proceedings of the XVI International Conference on High Energy Physics*, Chicago–Batavia, IL, J. D. Jackson and A. Roberts, eds. (National Accelerator Laboratory, Batavia), Vol. 4, pp. 189-247. The remarks on charge-exchange corrections are on pp. 205-206. [68]

Poggio, E. C., H. R. Quinn, and S. Weinberg (1976). Smearing Method in the Quark Model. *Phys. Rev. D* **13**, 1958-1968. [67]

Pratap, M., J. Smith, and Z. E. S. Uy (1972). Total Cross Sections for the Reactions $2\gamma \to 3\pi$ and $e^+e^- \to e^+e^-3\pi$. *Phys. Rev. D* **5**, 269-270. [66]

6. Gravitation

Introduction

During the first half of the 1970's, I started to get interested in learning more about gravitational physics. When I was a graduate student at Princeton in the early 1960's, particle physics and gravitational physics were quite separate subjects, with the former the domain of Goldberger and Treiman, and the latter the domain of Wheeler and Dicke, to mention just a few key faculty members. Under the unstructured system at Princeton, I never took a course in gravitation, and for my general exam got by with the introduction to general relativity that I obtained by reading the text of Peter Bergmann (1942), as well as reading some of the original Einstein papers reprinted in a Dover edition. (Working through the Dover volume was a project of an informal reading and discussion group during my senior year at Harvard, organized by Norval Fortson, an experimental physics graduate student affiliated with the residential house where I lived then.) However, in the 1970's it became clear both that many new results had been obtained in general relativity, so that my undergraduate knowledge was out-of-date, and that general relativity was becoming part of the essential tool kit of people working in quantum field theory. Among the things that convinced me of this were reading the thesis of Stephen Fulling (1972) on scalar quantum field theory in de Sitter space, while I was working on the $O(5)$ formulation of QED, the work of 't Hooft and Veltman (1974) and Deser (1975) on one-loop divergences of quantum gravity, and the availability of the new books on gravitation of Weinberg (1972), Misner, Thorne, and Wheeler (1973), and Hawking and Ellis (1973).

My intention in writing my comprehensive Hawaii lectures in the summer of 1975 was to wind up my involvement with neutrino physics, so that I could turn to something new. Since in 1976 I was due for a sabbatical, and my family did not want to travel away from Princeton, I decided that to learn relativity I would take a "reverse sabbatical", by going to Princeton University to teach the relativity course for a year. So I spent my evenings during the 1975-1976 academic year reading the texts of Weinberg and of Misner, Thorne, and Wheeler, and then took my sabbatical during the 1976-1977 academic year, teaching both the fall term course in Special Relativity and the spring term continuation course in General Relativity. I also was the faculty advisor for John David Crawford, who did a senior thesis on experimental

tests for curvature squared additions to the gravitational action. With this reading and teaching as background, I embarked on a number of relativity-related research projects, described in the next two sections.

First Papers

My first papers on gravity were the working out of a very speculative idea, that gravitation might be a composite phenomenon, with the gravitational fields arising as composite "pairing" amplitudes of photons in analogy with the energy gap order parameter for superconductivity. In the paper Adler, Lieberman, Ng, and Tsao (1976), we looked for weak coupling singularities in the electromagnetic photon-photon ladder graph sum in a conformally flat spacetime, and found some resemblances to the helicity structure of graviton exchange amplitudes. In a follow-up paper (Adler, 1976) I gave a linearized Hartree formulation for the photon pairing problem in a general background metric. I was never able to establish a detailed connection between photon pairing amplitudes and graviton couplings in the general case, and the fact that no weak coupling singularities occurred in flat spacetime meant that one could not establish a connection with the standard results of linearized general relativity. In retrospect, the absence of pairing effects in flat spacetime could have been expected from a subsequent theorem of Weinberg and Witten (1980), that ruled out spin-2 composites under quite general assumptions, and effectively doomed the program as set up in the 1976 papers. However, a useful outcome of writing these papers was that it started me thinking more generally about the idea of gravitation as an effective theory, and in particular about Sakharov's ideas on gravitation, which I briefly discussed in the paper Adler (1976); following up this direction later on led to my work on the Einstein action as a symmetry-breaking effect, discussed in the next section.

A second topic that I worked on in 1976 was the regularization of the stress-energy tensor for particles propagating in a general background metric. In the paper Adler, Lieberman, and Ng (1977), we applied covariant point-splitting techniques to the Hadamard series for the Green's functions, which we used to define a regularized stress-energy tensor for vector and scalar particles. This was a very technical computation, and contained useful formulas among its results, but also produced an embarrassment: by our method of regularization, we did not find the trace anomaly that had been found by others using different methods. We rechecked our calculation carefully, but could not find the source of the discrepancy. The problem was finally resolved by Wald (1978) (in time to be described in a note added in proof to our 1977 paper). Wald had earlier (Wald, 1977) set up a general axiomatization for the stress-energy tensor, and in Wald (1978) had shown that it leads to an essentially unique result. Applying a point-separation method similar to ours, he had also found no trace anomaly, but then went on to note that there was a subtle error

in our analysis. We had assumed that the local and boundary-condition-dependent parts of the Hadamard solution are separately symmetric in their arguments, but this is in fact not the case; only their sum is symmetric. Wald (1978) showed in the scalar case that when the analysis is repeated without the incorrect assumption, one gets the standard trace anomaly. Judy Lieberman and I then did the corresponding calculation in the vector particle case (Adler and Lieberman, 1978, R40), again finding that when the asymmetry of the two pieces of the Hadamard solution is taken into account, one gets the correct trace anomaly.

In a lunchtime conversation at some point during the 1977-1978 academic year, Robert Pearson asked whether the "no-hair" theorems of general relativity applied to the case of spontaneous symmetry breaking. I thought this was interesting and looked into it, finding no relevant papers in the literature. This became the subject of a joint paper (Adler and Pearson, 1978, R41), which showed that the standard "no-hair" theorems generalize to the vector field in the Abelian Higgs model, and to the non-conformally invariant Goldstone scalar field model. In our paper, we restricted ourselves to static, spherically symmetric black holes, and made the physically motivated assumption that any "hair" would also be static and spherically symmetric. This permits a simplifying choice of gauge for the Abelian Higgs model introduced by Bekenstein (1972). He observes that static electric charge distributions must give rise to static electric fields and vanishing magnetic fields. Thus one can find a special gauge in which the potentials A_μ obey $\vec{A} = 0$, $dA_0/dt = 0$. Since the gauge-independent source current j_μ obeys similar conditions $\vec{j} = 0$, $dj_0/dt = 0$, and since the gauge-independent magnitude of the Higgs scalar field is static, one finds that the residual phase of the Higgs scalar field in the special gauge is a space-independent, linear function of time, which can be eliminated by a further gauge transformation that preserves the gauge conditions $\vec{A} = 0$, $dA_0/dt = 0$. Thus one can do the analysis of possible "hair" taking the vector potential to be zero, and the Abelian Higgs field to be real. I have described Bekenstein's argument here in some detail because the choice of gauge in R41 is the basis of rather loosely worded objections to our paper in lectures of Gibbons (1990); his assertion (and that of authors who have quoted his lectures) that the gauge choice is problematic is not correct, as working through the Bekenstein argument given above makes clear. Also, I have rechecked the proof given in R41, and apart from the minor problem found by Ray, as discussed below, I find that the proof is correct, in disagreement with further statements in Gibbons' lecture. However, in response to Gibbons' comments about our choice of gauge, proofs of the "no-hair" theorem for the Abelian Higgs model that do not use a special gauge choice have since been given by Lahiri (1993) and by Ayón-Beato (2000).

Our argument starting from Eq. (24) of R41 was subsequently considerably simplified, and in the case when $d\theta/d\lambda|_H = 0$ corrected, in a paper of Ray (1979). (The subscript H here refers to evaluation at the horizon; see R41 for details of this and

other notation used in the following discussion of Ray's paper.) The minor problem noted by Ray resulted from our not dropping the subdominant term $d\theta/d\lambda$ on the right-hand side of Eq. (31) when integrating this equation to get Eq. (33), so as to be consistent with our dropping this term elsewhere, such as in Eq. (32). When this term is dropped, the $\theta^{-1/2}$ factor in Eq. (33) is replaced by a constant, and the approximate solution of Eq. (33) agrees with the exact solution of Eq. (24) given by Ray. As Ray points out, with this correction one still finds that $q^{-1}\phi^2$ is infinite at the horizon unless $K = 0$, which is what is needed to complete the proof.

Finally, I note that the subject of black hole "hair" in gauge theories has taken on new interest recently with the discovery that topological charges on a black hole can give nonzero effects outside the horizon; see, for example, Coleman, Preskill, and Wilczek (1992) and the related lectures of Wilczek (1998).

Einstein Gravity as a Symmetry Breaking Effect

In late January of 1978 I organized a small conference on "Geometry, Gravity and Field Theory" for the EST Foundation in San Francisco; this was a memorable event that was attended by a large fraction of the leading people with interests in quantum gravity. During my plane travel for this conference, and afterwards, I started to think about the confinement problem in QCD, and this became the main focus of my research for the next two years, as described in the following chapter. However, learning about scale breaking in QCD also led me back into gravitational physics, through considering the role similar mechanisms might play in giving a quantitative form to the suggestion by Sakharov (1968) (see also Klein, 1974) that Einstein gravity is the "metric elasticity" of spacetime. I did not arrive at the correct formulation immediately; I find in my files two unpublished manuscripts, the first positing monopole boundary conditions, and the second positing dimension-2 operators, as a source for symmetry breaking, in both cases suggesting connections with the Einstein-Hilbert action. I went as far as submitting a manuscript based on the second for publication, and also gave a seminar on it at Princeton University, where my arguments were torn to shreds by David Gross (following which I withdrew the manuscript). The criticism proved useful; I went home, learned more about dimensional transmutation and the theory of calculability versus renormalizability, and came up with the correct formulation given in Adler (1980a), R42. The basic idea here is that in theories which contain no scalars, so that scale invariance is spontaneously broken (QCD is a prime example, but "technicolor" type unification models also fit this description), there will be an induced order R term in the action in a curved background, with a coefficient that is calculable in terms of the scale mass of the theory. Thus, if an underlying unified theory spontaneously breaks scale invariance at the Planck scale, one can induce the Einstein gravitational action as a

scale-symmetry breaking effect, giving an explicit realization of the Sakharov–Klein idea.

I followed up this paper with a second one (Adler, 1980b, R43) in which I gave an explicit formula for the "induced gravitational constant" in theories with dynamical breakdown of scale invariance, expressed in terms of the vacuum expectation of the autocorrelation function of the trace of the renormalized stress-energy tensor $\tilde{T}_{\mu\nu}$,

$$(16\pi G_{\text{ind}})^{-1} = \frac{i}{96}\int d^4x[(x^0)^2 - (\vec{x})^2]\langle T(\tilde{T}^{\lambda}_{\lambda}(x)\tilde{T}^{\mu}_{\mu}(0))\rangle^{\text{flat spacetime}}_{0,\,\text{connected}} .$$

This formula for the induced Newton constant was independently obtained at about the same time by Zee (1981), and in the subsequent literature, the term "induced gravity" has come to be frequently used to describe the whole set of ideas involved. These papers attracted considerable attention in the gravity community, one result of which was that Claudio Teitelboim and his colleagues at the University of Texas in Austin invited me to give the Schild lectures in April of 1981. (My four lectures over a two week period, entitled "Einstein Gravity as a Symmetry-Breaking Effect in Quantum Field Theory", were the eleventh in the Schild series.) This proved memorable for an unanticipated reason; shortly before I was to go to Texas I contracted a mild case of what was probably type-A hepatitis (the kind transmitted by shellfish), and so was sick in bed with very little energy. I dragged myself out of bed on alternate days to write lecture notes, and then was so tired I had to sleep the entire day following. At any rate, I improved enough so that my doctor gave me permission to go to Texas, where Philip Candelas took me into his home and helped me get through my scheduled lectures. Ultimately, I expanded the lectures into a much-cited comprehensive article that appeared in *Reviews of Modern Physics* (Adler, 1982, R44). A year later, I wrote a briefer synopsis of the program of generating the Einstein action as an effective field theory, for a Royal Society conference on "The Constants of Physics", which was published as Adler (1983).

The explicit formula for the induced gravitational constant raises a number of interesting issues. First of all, if one assumes an unsubtracted dispersion relation for the Fourier transform $\psi(q^2)$ of the autocorrelation function of the stress-energy tensor trace, the induced gravitational constant is negative. However, as shown by Khuri (1982a) using analyticity methods, in asymptotically free theories there are three possible cases, depending on the distribution of zeros of $\psi(q^2)$, and in one of these cases G_{ind} has positive sign. In further papers Khuri (1982b,c) showed that in this case one can place useful bounds on the induced gravitational constant, expressed in terms of the scale mass of the theory.

The question of whether the formula for the induced gravitational constant gives a unique answer has been discussed, from the point of view infrared renormalon singularities, by David (1984) and in a follow-up paper of David and Strominger (1984). These authors argue that renormalons introduce an arbitrariness into the calculation

of $G_{\rm ind}$, as manifested through the fact that in the dimensional regularization of the ultraviolet singular "comparison function" $\Psi_c(t)$ introduced in Eq. (5.48) of R44, one has to continue onto a cut. In Appendix B, Section 3 of R44, I used a principal value prescription to deal with this, which David argues can be modified by taking complex weightings of the upper and lower sides of the branch cut, allowing a free parameter multiple of the imaginary part to be introduced into the calculation of the integral over the comparison function. David argues that this means that the expression for $G_{\rm ind}$ has an inherent ambiguity. I believe that this conclusion is suspect; since QCD and similar theories that spontaneously generate a mass scale are believed to be consistent field theories, their curved spacetime embeddings should, by the equivalence principle, also be consistent theories. This strongly suggests that the coefficient of the order R term in a curvature expansion of the vacuum action functional should be well defined, and that the ambiguity is an artifact of the comparison function procedure. This view is supported by the review article of Beneke (1999) on renormalons, where it is argued that renormalon ambiguities are typically canceled by corresponding ambiguities in non-perturbative terms (such as the integral ΔI_{UV} with integrand $\Psi - \Psi_c(t)$ in Eq. (5.48)), giving total physical amplitudes that are unambiguous. In other words, the renormalon ambiguities are an artifact of an attempted separation of QCD physical amplitudes into a "perturbative" and a "non-perturbative" part, and only indicate that if a branching prescription (such as a principal value) is needed for the perturbative part, then a corresponding branching prescription is also needed for the non-perturbative part. This will make the calculation of quantities like $G_{\rm ind}$ difficult, but does not imply that the calculation cannot, in principle, give a unique, physical answer. In the paper of David and Strominger (1984), the authors show that $G_{\rm ind}$ is unambiguous in *finite* supersymmetric theories, giving an existence proof that there are theories with a finite induced Newton's constant. In the general case, they acknowledge that "there is no *proof* that $G_{\rm ind}$ will *necessarily* be ambiguous", and I suspect that in fact $G_{\rm ind}$ will turn out to be well defined in a much wider class of supersymmetric and non-supersymmetric theories than only finite ones. Clearly, this is a question that merits further study.

If one thinks more generally about the structure of a fundamental theory of gravitation, there are a number of possibilities. It may be that the Planck length is the minimum length scale possible, because of an underlying "graininess" of spacetime. Or spacetime may be a continuum, as generally assumed, in which case the Planck length plays the role of the scale at which a classical metric breaks down, with new dynamical principles taking over at shorter distances. The suggestion that the order R gravitational action is an expression of scale symmetry breaking in a more fundamental scale-invariant theory is clearly based on a continuum picture of spacetime. A continuum assumption is also made in string theories, which however are not scale-invariant; in string theories a fundamental length scale (the string tension) appears in the action, and this directly sets the scale for the gravitational

action. Should spacetime turn out to be discrete or grainy, there may be more general forms of the induced gravitation idea that are relevant. Ultimately, the origin of the spacetime metric, and of the Einstein–Hilbert gravitational action that governs its dynamics, will not be certain until we have a unifying theory that also resolves the cosmological constant problem, which is not dealt with in any of the current ideas about quantum gravity.

References for Chapter 6

Adler, S. L. (1976). Linearized Hartree Formulation of the Photon Pairing Problem. *Phys. Rev. D* **14**, 379-383. [75]

Adler, S. L. (1980a) R42. Order-R Vacuum Action Functional in Scalar-Free Unified Theories with Spontaneous Scale Breaking. *Phys. Rev. Lett.* **44**, 1567-1569. [77]

Adler, S. L. (1980b) R43. A Formula for the Induced Gravitational Constant. *Phys. Lett. B* **95**, 241-243. [78]

Adler, S. L. (1982) R44. Einstein Gravity as a Symmetry-Breaking Effect in Quantum Field Theory. *Rev. Mod. Phys.* **54**, 729-766. [78,79]

Adler, S L. (1983). Einstein Gravitation as a Long Wavelength Effective Field Theory. *Phil. Trans. R. Soc. Lond. A* **310**, 273-278. This paper also appears in *The Constants of Physics*, W. H. McCrea and M. J. Rees, eds. (The Royal Society, London, 1983), pp. [63]-[68]. [78]

Adler, S. L. and J. Lieberman (1978) R40. Trace Anomaly of the Stress-Energy Tensor for Massless Vector Particles Propagating in a General Background Metric. *Ann. Phys.* **113**, 294-303. [76]

Adler, S. L., J. Lieberman, and Y. J. Ng (1977). Regularization of the Stress-Energy Tensor for Vector and Scalar Particles Propagating in a General Background Metric. *Ann. Phys.* **106**, 279-321. [75,76]

Adler, S. L., J. Lieberman, Y. J. Ng, and H.-S. Tsao (1976). Photon Pairing Instabilities: A Microscopic Origin for Gravitation? *Phys. Rev. D* **14**, 359-378. [75]

Adler, S. L. and R. B. Pearson (1978) R41. "No-Hair" Theorems for the Abelian Higgs and Goldstone Models. *Phys. Rev. D* **18**, 2798-2803. [76,77]

Ayón-Beato, E. (2000). "No-Hair" Theorem for Spontaneously Broken Abelian Models in Static Black Holes. *Phys. Rev. D* **62**, 104004. [76]

Bekenstein, J. (1972). Nonexistence of Baryon Number for Static Black Holes. *Phys. Rev. D* **5**, 1239-1246. [76]

Beneke, M. (1999). Renormalons. *Physics Reports* **317**, 1-142. See especially Sec. 2. [67,79]

Bergmann, P. G. (1942). *Introduction to the Theory of Relativity* (Prentice Hall, Englewood Cliffs). [74]

Coleman, S., J. Preskill, and F. Wilczek (1992). Quantum Hair on Black Holes. *Nucl. Phys. B* **378**, 175-246. [77]

David, F. (1984). A Comment on Induced Gravity. *Phys. Lett. B* **138**, 383-385. [78,79]

David, F. and A. Strominger (1984). On the Calculability of Newton's Constant and the Renormalizability of Scale Invariant Quantum Gravity. *Phys. Lett. B* **143**, 125-129. [78,79]

Deser, S. (1975). Quantum Gravitation: Trees, Loops and Renormalization, in *Quantum Gravity, an Oxford Symposium*, C. J. Isham, R. Penrose, and D. W. Sciama, eds. (Clarendon Press, Oxford), pp. 136-173. [74]

Fulling, S. (1972). Scalar Quantum Field Theory in a Closed Universe of Constant Curvature. Princeton University Dissertation. [74]

Gibbons, G. W. (1990). Self-Gravitating Magnetic Monopoles, Global Monopoles and Black Holes, in *The Physical Universe: The Interface between Cosmology, Astrophysics, and Particle Physics, Lecture Notes in Physics Vol. 383*, J. D. Barrow, A. B. Henriques, M. T. V. T. Lago, and M. S. Longair, eds. (Springer-Verlag, Berlin, 1991), pp. 110-133. [76]

Hawking, S. W. and G. F. R. Ellis (1973). *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge). [74]

Khuri, N. N. (1982a). Sign of the Induced Gravitational Constant. *Phys. Rev. D* **26**, 2664-2670. [78]

Khuri, N. N. (1982b). Upper Bound for Induced Gravitation. *Phys. Rev. Lett.* **49**, 513-516. [78]

Khuri, N. N. (1982c). Induced Gravity and Planck Zeros. *Phys. Rev. D* **26**, 2671-2680. [78]

Klein, O. (1974). Generalization of Einstein's Principle of Equivalence so as to Embrace the Field Equations of Gravitation. *Phys. Scr.* **9**, 69-72. [77,78]

Lahiri, A. (1993). The No-Hair Theorem for the Abelian Higgs Model. *Mod. Phys. Lett. A* **8**, 1549-1556. [76]

Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973). *Gravitation* (W. H. Freeman, San Francisco). [74]

Ray, D. (1979). Comment on the "No-Hair" Theorem for the Abelian-Higgs Model. *Phys. Rev. D* **20**, 3431. [76,77]

Sakharov, A. D. (1968). Vacuum Quantum Fluctuations in Curved Space and the Theory of Gravitation. *Dok. Akad. Nauk. SSSR* **177**, 70-71 (English translation: *Soviet Phys. - Doklady* **12**, 1040-1041). [77,78]

't Hooft, G. and M. Veltman (1974). One-Loop Divergencies in the Theory of Gravitation. *Ann. Inst. Henri Poincaré A: Physique théorique* **20**, 69-94. [74]

Wald, R. M. (1977). The Back Reaction Effect in Particle Creation in Curved Spacetime. *Commun. Math. Phys.* **54**, 1-19. [75]

Wald, R. M. (1978). Trace Anomaly of a Conformally Invariant Quantum Field in Curved Spacetime. *Phys. Rev. D* **17**, 1477-1484. [75,76]

Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (John Wiley, New York). [74]

Weinberg, S. and E. Witten (1980). Limits on Massless Particles. *Phys. Lett. B* **96**, 59-62. [75]

Wilczek, F. (1998). Lectures on Black Hole Quantum Mechanics: Lectures 3 & 4, in *The Black Hole 25 Years After*, C. Teitelboim and J. Zanelli, eds. (World Scientific, Singapore), pp. 229-326. [77]

Zee, A. (1981). Spontaneously Generated Gravity. *Phys. Rev. D* **23**, 858-866. [78]

7. Non-Abelian Monopoles, Confinement Models, and Chiral Symmetry Breaking

Introduction

The somewhat disparate topics to be discussed in this chapter are all connected through my interest during the late 1970's and early 1980's in studying nonperturbative properties of quantum chromodynamics (QCD), the theory of the strong interactions. I began these investigations by looking for a semi-classical model for heavy quark confinement. My first idea, that quarks might be confined in a non-Abelian monopole background field, did not work, but led to interesting progress in the theory of monopoles, as described in the first section. Most significantly, as discussed in detail, the monopole work led indirectly to the completion by Clifford Taubes of his multimonopole existence theorem during a visit to the IAS in the spring of 1980. I then turned to models based on the nonlinear dielectric properties of the QCD vacuum, which led to the confinement of quarks in "bag"-like structures which yield good heavy quark static potentials, as discussed in the second section. Finally, at the end of this period I worked briefly on the spontaneous breaking of chiral symmetry in QCD within the framework of pairing models patterned after superconductivity, as discussed in the final section. All three of these aspects of my study of QCD involved heavy numerical work, which in turn led to my interest in algorithms discussed in the next chapter.

Non-Abelian Monopoles

My first attempt at the confinement problem, which did not succeed but which had useful by-products that I shall describe here, was based on the idea of considering the potential between classical quark sources in the background of a non-Abelian 't Hooft (1974a)–Polyakov (1974)–Prasad–Sommerfield (1975)–Bogomol'nyi (1976) monopole or its generalizations, which I conjectured in Adler (1978b), R45 might act as a quark-confining "bag". To justify considering classical quark sources, I initially resorted to a scheme (Adler, 1978a) that I called "algebraic chromodynamics", which involved looking at the color space spanning the direct product of independent color charge matrices. However, I eventually dropped this apparatus in my pursuit of the confinement problem, and used instead the popular approximation of color charge matrices lying in a maximal Abelian subgroup of the $SU(3)$ color group of

conventional QCD, which gives a good first approximation to the full QCD color structure. Since it is clear that source charges in classical Yang–Mills theory are not confined, I looked for a simple modification of this theory that might lead to a linear potential. The first idea I tried was to look at classical Yang–Mills charges in the field of a background monopole. This had the obvious problem that the monopole scale has no clear relation to the QCD scale set by dimensional transmutation, but I simply ignored this difficulty and plunged ahead.

To pursue (and ultimately rule out) the conjecture that a monopole background would confine, I did a number of calculations of properties of monopole solutions. The first was a calculation of the Green's function for a single Prasad–Sommerfield monopole, by using the multi-instanton representation of the monopole and a formalism for calculating multi-instanton Green's functions given by Brown et al. (1978). This calculation was spread over two papers that I wrote; setting up contour integral expressions for the Green's function was done in Appendix A of Adler (1978b), R45, and the final result for the monopole propagator, after evaluation of the contour integrals and considerable algebraic simplification, was given in Appendix A of Adler (1979a), R46. (The fact that many lengthy expressions for parts of the Green's function collapsed, after algebraic rearrangement, into simple formulas, suggested that there should be a more efficient way to find the monopole Green's function. Not long afterwards, Nahm (1980) gave a new representation for the monopole that permitted a much simpler calculation of the Green's function given in R46.) To check that the lengthy expression that I had obtained for the propagator really satisfied the differential equation for the Green's function, I used numerical methods, calculating the partial derivatives acting on the propagator by finite difference methods on a very fine mesh. From numerical calculations based on the propagator formula, it was clear that a single monopole background would not lead to confinement; all that happened was that a Coulombic attractive $-1/r$ potential was reversed into a repulsive $1/r$ potential for large quark separations, a result that could have been anticipated from the large distance structure of the monopole field.

Not yet ready to give up on the monopole background idea, I then wrote two papers speculating that the Prasad–Sommerfield monopole might be a member of a larger class of solutions, in which the point at which the monopole Higgs field vanishes is extended to a higher-dimensional region, and in particular to a "string"-like configuration with a line segment as a zero set. In the first of these papers (Adler 1979b) I studied small deformations around the Prasad–Sommerfield monopole and found several series of such deformations. For normalized deformations I recovered the monopole zero modes that had already been obtained by Mottola (1978, 1979), but I found that "if an axially symmetric extension exists, it cannot be reached by integration out along a tangent vector defined by a nonvanishing, non-singular small-perturbation mode". This work was later extended into a complete calculation of the perturbations around the Prasad–Sommerfield solution by Akhoury, Jun, and

Goldhaber (1980), who also found "no acceptable nontrivial zero energy modes." In my second paper, Adler (1979c), I employed nonperturbative methods and suggested that despite the negative perturbative results, there might still be interesting extensions of the Prasad–Sommerfield solution with extended Higgs field zero sets.

At just around the same time, Erick Weinberg wrote a paper (Weinberg, 1979b) extending an index theorem of Callias (1978) to give a parameter counting theorem for multi-monopole solutions. Weinberg concluded that "any solution with n units of magnetic charge belongs to a $(4n-1)$-parameter family of solutions. It is conjectured that these parameters correspond to the positions and relative $U(1)$ orientations of n noninteracting unit monopoles". For $n = 1$, his results agreed with the zero-mode counting implied by Mottola's explicit calculation. Weinberg and I were aware of each other's work, as evidenced by correspondence in my file dating from March to June of 1979, and references relating to this correspondence in our papers Adler (1979c) and Weinberg (1979b).

My contact with Clifford Taubes was initiated by an April, 1979 letter from Arthur Jaffe, after I gave a talk at Harvard while Jaffe, as it happened, was visiting Princeton! In his letter, Jaffe noted that I was working on problems similar to those on which his students were working, and enclosed a copy of a paper by Clifford Taubes. (This preprint was not filed with Jaffe's letter, so I am not sure which of the early Taubes papers listed on the SLAC Spires archive that it was.) Jaffe's letter initiated telephone contacts with Taubes and some correspondence from him. On Jan. 6, 1980 Taubes wrote to me that he was making progress in proving the existence of multi-monopole Prasad–Sommerfield solutions, and in this letter and a second one dated on January 18, 1980 he reported results that were relevant to my conjectures on the possibility of deformed monopoles. His results placed significant restrictions on my conjectures; in a letter dated Feb. 1, 1980 I wrote to Lochlainn O'Raifeartaigh, who had also been interested in axially symmetric monopoles, saying that "On thinking some more about your paper (O'Raifeartaigh's preprint was unfortunately not retained in my files) I realized that the enclosed argument by Cliff Taubes is strong evidence against $n = 2$ monopoles involving a line zero. What Taubes shows is that a finite action solution of the Yang–Mills–Higgs Lagrangian cannot have a line zero of arbitrarily great length; hence if $n = 2$ monopoles contained a line zero joining the monopole centers, the monopole separation would be bounded from above. But this seems unlikely...". This correspondence and the result of Taubes was mentioned at the end of the published version, Houston and O'Raifeartaigh (1980).

As a result of our overlapping interests, I arranged for Taubes to make an informal visit, of two or three months, to the IAS during the spring of 1980. Clifford had expressed interest in this, he noted in a recent email, in part because Raoul Bott had suggested that he visit the Institute to get acquainted with Karen Uhlenbeck, who was visiting the IAS that year. In the course of his visit he met and interacted

with Uhlenbeck, who, along with Bott, had a major impact on his development as a mathematician.

Taubes began the visit by looking at my conjecture of extended zero sets, but after a while told me that he could not find an argument for them. Partly as a result of his work, I was getting disillusioned with my own conjecture, so I asked him what was happening with his attempted proof of multi-monopole solutions. Taubes replied that he was stuck on that, and not sure whether they existed. I then mentioned to him Erick Weinberg's parameter counting result, which strongly suggested a space of moduli much like that in the instanton case, where looking at deformations correctly suggests the existence and structure of the multi-instanton solutions. To my surprise, Taubes was not aware of Erick's result, and knowing it impelled him into action on his multi-monopole proof. Within a week or two he had completed a proof, and wrote it up on his return to Harvard. (Thus, there was a parallel to what happened a year before with respect to solutions of the first order Ginzburg–Landau equations. In that case, Taubes had heard a lecture at Harvard by Erick Weinberg on parameter counting for multi-vortex solutions (written up as Weinberg, 1979a) and then went home and came up with his existence proof for multi-vortices, Taubes (1980). The vortex work provided the initial impetus for Taubes' turning to the monopole problem.) In his paper Taubes showed that "for every integer $N \neq 0$ there is at least a countably infinite set of solutions to the static $SU(2)$ Yang-Mills-Higgs equations in the Prasad-Sommerfield limit with monopole number N. The solutions are partially parameterized by an infinite sublattice in $S_N(R^3)$, the N-fold symmetric product of R^3 and correspond to noninteracting, distinct monopoles." This quote is taken from the Abstract of his preprint "The Existence of Multi-Monopole Solutions to the Static, $SU(2)$ Yang–Mills–Higgs Equations in the Prasad–Sommerfield Limit", which was received on the SLAC Spires data base in June, 1980, and which carried an acknowledgement on the title page noting that "This work was completed while the author was a guest at the Institute for Advanced Studies, Princeton, NJ 08540". His preprint also ended with an Acknowledgment section noting his conversations with me, with Arthur Jaffe, and with Karen Uhlenbeck, as well as the Institute's hospitality. The proof was not published in this form, however, but instead appeared (with acknowledgments edited out at some stage) as Chapter IV of the book Jaffe and Taubes (1980) that was completed in August of 1980. The multimonopole existence proof was a milestone in Taubes' career; in a recent exchange of emails relating to the events described in this section, Taubes commented on his visit "to hang out at the IAS during the spring of 1980. It profoundly affected my subsequent career...". He went on to further investigations of monopole solutions, that lead him to studies of 4-manifold theory which have had a great impact on mathematics.

O'Raifeartaigh, who had been following the monopole work at a distance, invited me during the spring of 1980 to come to Dublin that summer to lecture on my papers.

However, since Taubes had much more interesting results I suggested to Lochlainn that he ask Clifford instead, and Taubes did go to Dublin to lecture. After Clifford's visit, I redirected my search for semiclassical confinement models to a study of nonlinear dielectric models by analytic and numerical methods, in collaboration with Tsvi Piran; these models do give an interesting class of confining theories, and are described in the following section. Based on the observation that the Yang–Mills action is multiquadratic (that is, at most quadratic in each individual potential component), Piran and I also applied the same numerical relaxation methods to give an efficient method for the computation of axially symmetric multimonopole solutions. (This was done mainly to illustrate the computer methods, since by then exact analytic 2-monopole solutions had appeared; see Forgacs, Horvath and Palla (1981) and Ward (1981).) The numerical methods that Piran and I developed were described in our *Reviews of Modern Physics* article Adler and Piran (1984), R47 that marked the completion of the research program on confining models, and as a by-product, on monopoles.

Confinement Models

Having seen that monopole backgrounds would not confine, I turned my attention to another type of semi-classical model, proposed in various forms by Savvidy (1977) (see also Matinyan and Savvidy (1978)) and Pagels and Tomboulis (1978). The basic idea is to do electrostatics with Abelianized quark charges, and with the fundamental QCD action replaced by a renormalization group improved effective action, in which the gauge coupling is replaced by a running coupling, that is taken to be a function solely of the field strength squared. Although use of the running coupling is only justified by the renormalization group in the ultraviolet regime of large field strengths, the model assumes that the same functional form can be extrapolated to small field strengths as well. This leads to electrodynamics with a nonlinear, field-dependent dielectric constant that develops a zero for small squared field strengths. Because the only dynamical input from QCD is the running coupling, the model, as Frank Wilczek later remarked to me, can be considered as a very simple embodiment of the idea that "asymptotic freedom" should be associated with "infrared slavery". Since the running coupling involves a scale mass, the model directly incorporates the phenomenon of dimensional transmutation. Pagels and Tomboulis conjectured, on the basis of various evidence, that the nonlinear dielectric model would confine, but did not have a proof.

In the paper Adler (1981), R48, I analyzed the effective action model in detail and proved that it confines quarks. The argument starts from a Euclidean form of the Feynman path integral, and shows that the static potential is the minimum of the effective action in the presence of sources. I then specialized to the leading-logarithm effective action, and showed that the action minimum is associated with a

field configuration in which a color magnetic field fills in whenever the color electric field is less than the minimum magnitude κ at which the effective action is minimized. This reduces the action minimization to an electrostatics problem, to which one can apply flux conversation estimates due to 't Hooft (1974b). In the nonlinear dielectric model context, these estimates show that the static potential is bounded from below by $\kappa Q(R-r)$, with Q the Abelianized quark charge, with r a constant, and with R the interquark separation. Hence the potential increases linearly for large R, and the model confines. In an Appendix to R48, I discussed how a one-loop renormalization group exact, leading-logarithm running coupling can be obtained, by a coupling constant transformation, from the more usual two-loop renormalization group exact running coupling (to which the confinement argument also applies).

When I presented this proof of confinement by the nonlinear dielectric model at a Department of Energy sponsored workshop in Yerevan, Armenia in 1983, an interesting dialogue with the Soviet physicist A. B. Migdal ensued. When I started to talk, and said what I was going to prove, Migdal stood up and stated that it was well-known that the Savvidy (–Pagels–Tomboulis) model did not confine, and gave some reasons. I then presented my proof, after which Migdal stood up, and said words to the effect that the problem is that there are too many confining models! As we shall see, there is really only one other model, the "dual superconductor" model, which like the nonlinear dielectric model is motivated by the idea of a color magnetic condensate, but describes this with a different dynamics.

Following publication of R48, I wrote a paper (Adler, 1982a) formalizing the approximations (further discussed below) needed to get an Abelianized effective action model from the functional integral for QCD. I then turned to the problem of understanding in detail *how* the leading-log model gives a confining potential. Since it was clear that this would, at least in part, involve numerical solution of the nonlinear differential equations involved, I brought in Tsvi Piran as a postdoc. Tsvi had worked extensively in the numerical solution of the Einstein equations of general relativity, and came to the IAS with the understanding that he would continue this and other interests he had in astrophysics, but would also collaborate with me in the numerical solution of the leading-log model equations. Because of my work on the induced gravity program, this collaboration didn't start immediately after Tsvi's arrival, but once we began work, Tsvi taught me a great deal about setting up an interactive program to numerically solve partial differential equations. As is typical in doing numerical work, most of our time was spent developing and testing our computer codes, which took many months. To guard against programming errors, Tsvi and I each independently wrote our own programs, which once debugged gave identical results. The final production runs took a total of less than two days running time on the then new IAS VAX 11/780 computer, using mesh sizes of up to 100×100 to resolve details of the confinement domain.

The equations to be solved, in the leading-log model with three light fermion

flavors and scale mass κ, are

$$\begin{aligned}
\vec{\nabla}\cdot(\epsilon(E)\vec{E}) &= j^0 \quad , \\
j^0 &= Q\delta(x)\delta(y)[\delta(z-a)-\delta(z+a)] \quad , \\
\epsilon(E) &= \frac{1}{4}b_0\log(E^2/\kappa^2) \quad , \quad E=|\vec{E}| \quad , \\
b_0 &= \frac{9}{8\pi^2} \quad .
\end{aligned}$$

We also studied the leading-log-log model, in which the two-loop exact form of the running coupling is used. We originally tried to solve the equations directly in terms of the scalar potential A^0, but found that the numerical programs were unstable. I then introduced a flux function reformulation of the problem (suggested by similar methods used in plasma physics), and this gave a stable, rapidly convergent iteration showing formation of a flux-confining free boundary. To understand the structure of the free boundary, Tsvi suggested that a paper of Fichera on elliptic equations that degenerate to parabolic would be relevant, and this indeed was the case, as described in Appendix A of our review R47. Prior to writing the review, we wrote two shorter papers. The first (Adler and Piran, 1982a, R49) demonstrated flux confinement and gave a numerical determination of the large R asymptotic form of the interquark potential, which contains a leading term linear in R, and a subdominant term proportional to $\log \kappa^{1/2}R$. The second (Adler and Piran, 1982b, R50) gave compact, accurate functional forms that fit the computed static potentials for both the leading-log and the leading-log-log models.

One nice feature of the leading-log model (as well as the leading-log-log extension) is that its small distance and large distance limiting cases can be approximated analytically. In the small distance limit, I devised an analytic perturbation method (Adler, 1982b) which shows that the potential has the standard form of a Coulomb potential with a logarithmic correction that is expected from perturbative QCD, permitting the parameter κ of the model to be related to the QCD scale mass. With this identification, the model has no adjustable parameters. In the large distance limit, an ingenious analysis by Lehmann and Wu (1984) showed that the confinement domain is an ellipsoid of revolution, with maximum diameter growing as $R^{1/2}$ with the interquark separation, and gave an analytic expression for the free boundary shape for large R as well as the subdominant term in the potential. Thus, the model yields a "fat" bag, rather than a cylindrical confinement domain of uniform radius; however, Lehmann told me at the time that he believed the true QCD behavior would show a constant-radius cylindrical domain, and he appears now (see below) to be right. As discussed in the articles I wrote with Piran, the analytic forms for both small and large R agreed very well with our numerical results, giving confidence that the numerical analysis had been carried out correctly.

How well do the nonlinear dielectric models agree with QCD? There are two

aspects to this question, whether they give satisfactory static potentials, and whether they describe the flux confinement domain that is realized in QCD. To assess the static potentials tabulated in R50, one has to do a detailed fit to heavy quark spectroscopic data. This was done in papers of Margolis, Mendel, and Trottier (1986) and of Crater and Van Alstine (1988), both of which concluded that the log-log model potential is in good agreement with experimental data on heavy quark systems, with reasonable values of the quark masses. The fit of Margolis, Mendel, and Trottier used a value of $\Lambda_{\bar{M}\bar{S}} = 0.270\,\mathrm{GeV}$, while that of Crater and Van Alstine used a value of $\Lambda_{\bar{M}\bar{S}} = 0.215\mathrm{GeV}$ (note that their Λ is the $\kappa^{1/2}$ of R50, which is related to $\Lambda_{\bar{M}\bar{S}}$ by $\Lambda_{\bar{M}\bar{S}} = 0.959\kappa^{1/2}$). These values of $\Lambda_{\bar{M}\bar{S}}$ are in reasonable accord with the value $\Lambda_{\bar{M}\bar{S}} = 0.218\,\mathrm{GeV}$ that Piran and I had quoted in R50, obtained by requiring the best fit of our potential to Martin's phenomenological potential for heavy quark systems. These values of $\Lambda_{\bar{M}\bar{S}}$ should be compared with the three light quark experimental value $\Lambda^{(3)}_{\bar{M}\bar{S}} \simeq 0.369\,\mathrm{GeV}$ (Hinchliffe, 2005). For a simple extrapolation from the asymptotically free regime to the confining regime of QCD, the nonlinear dielectric model does reasonably well in accounting for heavy quark spectroscopy.

As already noted, the confinement domain in the nonlinear dielectric models is an ellipsoid of revolution, with width increasing with the quark separation R. Let ρ be the cylindrical radial coordinate, and z the coordinate along the axis of the cylinder. On the medial plane $z = 0$, various quantities of interest can be computed in the large-R limit directly from the Lehmann–Wu asymptotic solution. The radius of the confinement domain on the medial plane is

$$\rho_m = R^{\frac{1}{2}} \left(\frac{2Q}{\pi b_0 \kappa}\right)^{\frac{1}{4}} \quad ,$$

and the value of $|\vec{D}|$ on the medial plane is

$$|\vec{D}| = \frac{1}{R}\left(\frac{2Qb_0\kappa}{\pi}\right)^{\frac{1}{2}} (1 - \rho^2/\rho_m^2) \quad ,$$

from which one can check that the flux integral gives $2\pi \int_0^{\rho_m} \rho d\rho |\vec{D}| = Q$. The profile of $|\vec{D}|$ is evidently parabolic, and scales with $\rho_m \propto R^{\frac{1}{2}}$.

To compare this "fat bag" confinement domain with QCD, one must rely on lattice simulations, since in real-world QCD, the confining flux tube breaks up through quark-antiquark pair formation before the asymptotic regime is reached. Assuming that the lattices used are large enough to accurately approximate the continuum theory, the data from simulations that have been carried out show a confinement domain of constant diameter in the limit of large R, as discussed and referenced in the book of Ripka (2004). The details of the simulated confinement domain favor the "dual superconductor" model, in which QCD is regarded as a dual of a Ginzburg–Landau superconductor, with magnetic monopole pairs replacing the Cooper pairs

of superconductivity. In this picture, in addition to the color fields, there is a dynamical variable corresponding to the monopole condensate. A numerical analysis of flux confinement in a dual superconductor, using the methods described in my review with Piran R47, has been given by Ball and Caticha (1988), who give plots of the confinement domain; for further details and references, see both Ripka (2004) and the review of Baker, Ball, and Zachariasen (1991). For appropriate values of the three dual superconductor model parameters (a magnetic charge g, which can be related to an effective QCD coupling $e_{\rm eff}$ by the Dirac quantization condition $g = 2\pi/e_{\rm eff}$, a scalar magnetic condensate mass m_H, and a gauge gluon mass m_V), good fits to the lattice simulations are obtained, and the dual superconductor model also gives a phenomenologically satisfactory static potential. (In a recent preprint, Haymaker and Matsuki (2005) argue that in lattice comparisons, the continuum m_V gives rise to two parameters that must be fitted, making four parameters in all including g.) However, since the dual superconductor gives a Coulomb potential at short distances, without logarithmic modifications, the dual superconductor parameters cannot be directly related to the QCD scale $\Lambda_{\bar{M}\bar{S}}$ as was possible for the scale parameter κ of the nonlinear dielectric model. As a limiting case, the dual superconductor model gives the standard bag model with a field discontinuity at the boundary; a numerical solution of this model is also discussed in Ball and Caticha (1988).

Although the nonlinear dielectric model and the dual superconductor model successfully describe important aspects of confinement in QCD, major steps would be needed to incorporate such classical action models into a *proof* of confinement. To do so one would have to prove that the true energy of a widely separated quark-antiquark pair in QCD is bounded from below by the energy calculated in one or the other of the two models. This would require achieving precise control over the qualitative approximations involved in the models, which include a mean-field approximation to the functional integral as discussed in Adler (1982a), the replacement of the exact QCD effective action by the model effective action, and replacement of the octet of color quark charges by Abelianized effective charges lying in the maximal commutative subgroup. Although, as I argued in the case of the nonlinear dielectric model in Adler (1982a), these simplifications of the full problem are plausible, replacing qualitative approximations by precise mathematical statements with error estimates will be no small task.

In any flux tube picture of confinement based on Abelianized charges, such as either the nonlinear dielectric model or the bag limit of the Ginzburg–Landau dual superconductor model, the string tension scales as the Abelianized quark charge, or as the square root of the corresponding Casimir. In a paper with Neuberger (Adler and Neuberger, 1983, R51), we pointed out that in the large-N_c limit of $SU(N_c)$ gauge theory, the string tension scales with the Casimir when changing from fundamental to adjoint representation quarks; hence to the extent that flux tube

models give a good description of confinement in $N_c = 3$ QCD, different confinement mechanisms appear to be at work in QCD and in its large N_c limit.

Chiral Symmetry Breaking

Not long after I had finished the review paper R47 with Piran summarizing our work on confinement models, Anne Davis suggested looking at another outstanding problem in QCD, that of chiral symmetry breaking. After studying the relevant literature (reviewed in the Introduction to our paper Adler and Davis (1984), R52), we decided to focus on setting up and solving a superconductor-like gap equation for fermion pairing in Coulomb gauge, systematically imposing the axial-vector current Ward identities to get the correct renormalization procedure. This method permitted us to study pairing using a Lorentz vector instantaneous confining potential with $V \propto r$, getting infrared-finite results for physical quantities without imposing *ad hoc* infrared cutoffs. The model gives spontaneous breaking of chiral symmetry, but with values of the quark condensate $\langle \bar{u}u \rangle$ and the pion decay constant f_π that are considerably too low when the phenomenological confining potential (or string tension) is used as input. Similar results were also found by a group at Orsay, and we learned later that the utility of the axial-vector Ward identities in deriving the gap equation had also been noted by Delbourgo and Scadron (see the reprinted papers R52 and R53 for references). Extensions of the model of R52 to the finite temperature case were later discussed by Davis and Matheson (1984), Alkhofer and Ammundsen (1987), and Klevansky and Lemmer (1987).

In a subsequent paper (Adler, 1986, R53) that I wrote for the Nambu Festschrift, I reviewed the work of various groups on gap equation models, and also noted a problem. In order for there to be no explicit breaking of chiral symmetry in the gap equation model, the instantaneous potential must be the time component of a Lorentz vector, so that it contains factors γ_0 that anticommute with γ_5. However, experimental data on heavy quark spectroscopy show that the confining part of the potential is predominantly Lorentz scalar, and using a Lorentz scalar potential in the gap equation model would lead to explicit violation of chiral symmetry, and therefore invalidate the model. This suggests that the approximations leading to the gap equation model are not valid for the confining part of the potential. In R53, I also gave equations that I had worked out for a retarded extension of the instantaneous potential model. The original intention had been for a graduate student in either Princeton or Cambridge to work on solving the extended model, but in view of the Lorentz structure problem this was not done (a covariant treatment of the gap equation model was later given by von Smekal, Amundsen, and Alkofer (1991)). For various proposals for addressing the Lorentz structure issue, see Lagaë (1992), Szczepaniak and Swanson (1997), and Bicudo and Marques (2004).

Shortly after the paper R52 was out, Cumrun Vafa, then a Princeton graduate

student, had a few conversations with me about his attempts to turn the Banks–Casher (1980) eigenvalue density criterion for chiral symmetry breaking into a proof that chiral symmetry breaking occurs in QCD. (For recent progress in applying the Banks–Casher criterion in the large-N_c limit, see Narayanan and Neuberger (2004).) I didn't have much in the way of concrete suggestions to offer, and Cumrun started also talking to Edward Witten, who very sagely suggested looking at a different problem, that of studying whether parity conservation can be spontaneously broken in QCD. This problem proved tractable, and their papers (Vafa and Witten, 1984a,b), proving that parity is not spontaneously broken in vector-like gauge theories (and similarly for the isospin and baryon number symmetries), became part of Vafa's thesis. The difference between the parity problem and the chiral symmetry problem can be understood by considering their respective order parameters. If parity is spontaneously broken, the pseudoscalar order parameter $\overline{u}i\gamma_5 u$ will receive a vacuum expectation. When the fermions are integrated out, one obtains a Lorentz invariant, parity-nonconserving operator functional X of the gluon fields that is real in Minkowski space, but picks up a factor of i when rotated to Euclidean space. This, together with positivity of the Euclidean space Dirac fermionic determinant in a vector-like theory, is the basis of the Vafa-Witten proof that adding a small multiple of X to the action cannot make the ground state energy lower. In the chiral symmetry problem, the relevant order parameter is the parity conserving but chiral symmetry breaking scalar operator $\overline{u}u$, which when the fermions are integrated out leads to a functional X' of the gluon fields that remains real when rotated to Euclidean space. Hence the Vafa-Witten argument suggests that the energy minimum may lie at a nonzero value of X', but such a local analysis cannot find the global minimum, and hence does not give a proof of chiral symmetry breaking. Rigorous lattice inequalities given by Weingarten (1983) give a proof of chiral symmetry breaking only when additional strong assumptions are made, including the existence of the continuum limit and the confinement of color, together with use of anomaly matching conditions.

Over twenty years later, the problem of proving the breakdown of chiral symmetry in QCD is still open, as is that of proving confinement. In fact, there is considerable evidence that chiral symmetry breaking and confinement in QCD are related phenomena. For example, lattice simulations such as D'Elia et al. (2004) show that the deconfining and chiral transitions coincide; gap equation models of the type studied in R52 find chiral symmetry breaking for a confining potential but not for a Coulomb potential, and lattice inequalities of the type studied by Weingarten also need confinement as an ingredient to show chiral symmetry breaking. Thus it appears that both of these outstanding problems in QCD are aspects of the larger problem of proving that QCD exists and has a mass gap, which is one of the seven Clay Foundation Millennium Problems in mathematics and mathematical physics. Perhaps in this century, with the added incentive of a $1 million reward,

rigorous proofs of confinement and chiral symmetry breaking in QCD will be found!

References for Chapter 7

Adler, S. L. (1978a). Classical Algebraic Chromodynamics. *Phys. Rev. D* **17**, 3212-3224. [83]

Adler, S. L. (1978b) R45. Theory of Static Quark Forces. *Phys. Rev. D* **18**, 411-434. Pages 424-429, containing Appendix A, are reprinted here. [83,84]

Adler, S. L. (1979a) R46. Classical Quark Statics. *Phys. Rev. D* **19**, 1168-1187. Pages 1182-1183, containing Appendix A, are reprinted here. [84]

Adler, S. L. (1979b). Small Deformations of the Prasad-Sommerfield Solution. *Phys. Rev. D* **19**, 2997-3007. [84]

Adler, S. L. (1979c). Global Structure of Static Euclidean $SU(2)$ Solutions. *Phys. Rev. D* **20**, 1386-1411. [85]

Adler, S. L. (1981) R48. Effective-Action Approach to Mean-Field Non-Abelian Statics, and a Model for Bag Formation. *Phys. Rev. D* **23**, 2905-2915. [87,88]

Adler, S. L. (1982a). Generalized Bag Models as Mean-Field Approximations to QCD. *Phys. Lett. B* **110**, 302-306. [88,91]

Adler, S. L. (1982b). Short-Distance Perturbation Theory for the Leading Logarithm Models. *Nucl Phys. B* **217**, 381-394. [89]

Adler, S. L. (1986) R53. Gap Equation Models for Chiral Symmetry Breaking. *Progr. Theor. Phys. Suppl.* **86**, 12-17. [92]

Adler, S. L. and A. C. Davis (1984) R52. Chiral Symmetry Breaking in Coulomb Gauge QCD. *Nucl. Phys. B* **244**, 469-491. [92,93]

Adler, S. L. and H. Neuberger (1983) R51. Quasi-Abelian versus Large-N_c Linear Confinement. *Phys. Rev. D* **27**, 1960-1961. [91,92]

Adler, S. L. and T. Piran (1982a) R49. Flux Confinement in the Leading Logarithm Model. *Phys. Lett. B* **113**, 405-410. [89]

Adler, S. L. and T. Piran (1982b) R50. The Heavy Quark Static Potential in the Leading Log and the Leading Log Log Models. *Phys. Lett. B* **117**, 91-95. [89,90]

Adler, S. L. and T. Piran (1984) R47. Relaxation Methods for Gauge Field Equilibrium Equations. *Rev. Mod. Phys.* **56**, 1-40. Pages 1-12, 18-21, and 30-38 are reprinted here. [87,89,91,92,98]

Akhoury, R., J.-H. Jun, and A. S. Goldhaber (1980). Linear Deformations of the Prasad-Sommerfield Monopole. *Phys. Rev. D* **21**, 454-465. [84,85]

Alkhofer, R. and P. A. Amundsen (1987). A Model for the Chiral Phase Transition in QCD. *Phys. Lett. B* **187**, 395-400. [92]

Baker, M., J. S. Ball, and F. Zachariasen (1991). Dual QCD: A Review. *Physics Reports* **209**, 73-127. [91]

Ball, J. S. and A. Caticha (1988). Superconductivity: A Testing Ground for Models of Confinement. *Phys. Rev. D* **37**, 524-535. [91]

Banks, T. and A. Casher (1980). Chiral Symmetry Breaking in Confining Theories. *Nucl. Phys. B* **169**, 103-125. [93]

Bicudo, P. and G. M. Marques (2004). Chiral Symmetry Breaking and Scalar String Confinement. *Phys. Rev. D* **70**, 094047. [92]

Bogomol'nyi, E. B. (1976). The Stability of Classical Solutions. *Yad. Fiz.* **24**, 861-870 (English translation: *Sov. J. Nucl. Phys.* **24**, 449-454). [83]

Brown, L. S., R. D. Carlitz, D. B. Creamer, and C. Lee (1978). Propagation Functions in Pseudoparticle Fields. *Phys. Rev. D* **17**, 1583-1597. [84]

Callias, C. (1978). Axial Anomalies and Index Theorems on Open Spaces. *Commun. Math. Phys.* **62**, 213-234. [85]

Crater, H. W. and P. Van Alstine (1988). Two-Body Dirac Equations for Meson Spectroscopy. *Phys. Rev. D* **37**, 1982-2000. [90]

Davis, A. C. and A. M. Matheson (1984). Chiral Symmetry Breaking at Finite Temperature in Coulomb Gauge QCD. *Nucl. Phys. B* **246**, 203-220. [92]

D'Elia, M., A. Di Giacomo, B. Lucini, G. Paffuti, and C. Pica (2004). Chiral Transition and Deconfinement in $N_f = 2$ QCD; arXiv:hep-lat/0408009. [93]

Forgacs, P., Z. Horvath, and L. Palla (1981). Exact Multi-Monopole Solutions in the Bogomolny-Prasad-Sommerfield Limit. *Phys. Lett. B* **99**, 232-236; erratum, *Phys. Lett. B* **101**, 457 (1981). [87]

Haymaker, R. W. and T. Matsuki (2005). Consistent Definitions of Flux and the Dual Superconductivity Parameters in $SU(2)$ Lattice Gauge Theory; arXiv:hep-lat/0505019. [91]

Hinchliffe, I. (2005). Private communication. For the underlying formulas relating the $\Lambda^{(n)}_{\overline{MS}}$ for different values of n, see Hinchliffe's QCD review in the 1998 Review of Particle Physics, *Eur. Phys. J. C* **3**, 1-794, pp. 81-89. [90]

Houston, P. and L. O'Raifeartaigh (1980). On the Zeros of the Higgs Field for Axially Symmetric Multi-Monopole Configurations. *Phys. Lett. B* **93**, 151-154. [85]

Jaffe, A. and C. Taubes (1980). *Vortices and Monopoles* (Birkhäuser, Boston). [86]

Klevansky, S. P. and R. H. Lemmer (1987). Chiral-Symmetry Breaking at Finite Temperatures. *Phys. Rev. D* **38**, 3559-3565. [92]

Lagaë, J.-F. (1992). Spectroscopy of Light-Quark Mesons and the Nature of the Long-Range $q - \bar{q}$ Interaction. *Phys. Rev. D* **45**, 317-327. [92]

Lehmann, H. and T. T. Wu (1984). Classical Models of Confinement. *Nucl. Phys. B* **237**, 205-225. [89,90]

Margolis, B., R. R. Mendel, and H. D. Trottier (1986). $Q - \bar{q}$ Mesons in the Leading Log and Leading Log-Log Models. *Phys. Rev. D* **33**, 2666-2673. [90]

Matinyan, S. G. and G. K. Savvidy (1978). Vacuum Polarization Induced by the Intense Gauge Field. *Nucl. Phys. B* **134**, 539-545. [87]

Mottola, E. (1978). Zero Modes of the 't Hooft–Polyakov Monopole. *Phys. Lett. B* **79**, 242-244. [84,85]

Mottola, E. (1979). Normalizable Solutions to the Dirac Equation in the Presence of a Magnetic Monopole. *Phys. Rev. D* **19**, 3170-3172. [84,85]

Nahm, W. (1980). A Simple Formalism for the BPS Monopole. *Phys. Lett. B* **90**, 413-414. [84]

Narayanan, R. and H. Neuberger (2004). Chiral Symmetry Breaking at Large N_c. *Nucl. Phys. B* **696**, 107-140. [93]

Pagels, H. and E. Tomboulis (1978). Vacuum of the Quantum Yang-Mills Theory and Magnetostatics. *Nucl. Phys. B* **143**, 485-502. [48,87]

Polyakov, A. M. (1974). Particle Spectrum in Quantum Field Theory. *ZhETF Pis. Red.* **20**, 430-433 (English translation: *JETP Lett.* **20**, 194-195). [83]

Prasad, M. K. and C. M. Sommerfield (1975). Exact Classical Solution for the 't Hooft Monopole and the Julia-Zee Dyon. *Phys. Rev. Lett.* **35**, 760-762. [83]

Ripka, G. (2004). *Dual Superconductor Models of Color Confinement.* Lecture Notes in Physics 639 (Springer, Berlin). [90,91]

Savvidy, G. K. (1977). Infrared Instability of the Vacuum State of Gauge Theories and Asymptotic Freedom. *Phys. Lett. B* **71**, 133-134. [87]

Szczepaniak, A. P. and E. S. Swanson (1997). On the Dirac Structure of Confinement. *Phys. Rev. D* **55**, 3987-3993. [92]

Taubes, C. H. (1980). Arbitrary N-Vortex Solutions to the First Order Ginzburg–Landau Equations. *Commun. Math. Phys.* **72**, 277-292. [86]

't Hooft, G. (1974a). Magnetic Monopoles in Unified Gauge Theories. *Nucl. Phys. B* **79**, 276-284 (1974). [83]

't Hooft, G. (1974b). Quarks and Gauge Fields, in *Recent Progress in Lagrangian Field Theory and Applications*, Proceedings of the Marseille Colloquium, 1974, C. P. Korthals-Altes, E. de Rafael, and R. Stora, eds. (Centre de Physique Théorique -CNRS, Universités d'Aix-Marseille I et II, 1975), pp. 58-67. [88]

Vafa, C. and E. Witten (1984a). Parity Conservation in Quantum Chromodynamics. *Phys. Rev. Lett.* **53**, 535-536. [93]

Vafa, C. and E. Witten (1984b). Restrictions on Symmetry Breaking in Vector-Like Gauge Theories. *Nucl. Phys. B* **234**, 173-188. [93]

von Smekal, L., P. A. Amundsen, and R. Alkofer (1991). A Covariant Model for Dynamical Chiral Symmetry Breaking in QCD. *Nucl. Phys. A* **529**, 633-652. [92]

Ward, R. S. (1981). A Yang-Mills-Higgs Monopole of Charge 2. *Commun. Math. Phys.* **79**, 317-325. [87]

Weinberg, E. J. (1979a). Multivortex Solutions of the Ginzburg–Landau Equations. *Phys. Rev. D* **19**, 3008-3012. [86]

Weinberg, E. J. (1979b). Parameter Counting for Multimonopole Solutions. *Phys. Rev. D* **20**, 936-944. [85,86]

Weingarten, D. (1983). Mass Inequalities for Quantum Chromodynamics. *Phys. Rev. Lett.* **51**, 1830-1833. [93]

8. Overrelaxation for Monte Carlo and Other Algorithms

Introduction

As I have already noted, the investigations described in the previous chapter all involved extensive computer work. This got me interested in the issue of algorithms more generally, and led to two distinct research directions in the years that followed. One involved generalizing the acceleration methods for solving partial differential equations to the related problem of Monte Carlo simulations, as discussed in the first section that follows. The second involved neural networks and pattern recognition, and led among other things to work on image normalization methods, described briefly in the second section of this chapter.

Overrelaxation to Accelerate Monte Carlo

In preparation for numerically solving the partial differential equations for the leading-log models, I did general reading on numerical methods for handling partial differential equations. This taught me about the critical slowing down problem – the fact that as one refines meshes to get more accurate numerical solutions, the rate of convergence of the iterations slows down. I also learned about various strategies devised for defeating critical slowing down, and in particular about the successive over-relaxation (SOR) modification of the standard Gauss–Seidel iteration. In a Gauss–Seidel iteration of a positive functional, one replaces each successive variable by the value that locally minimizes the functional. In SOR, one builds in a systematic overshoot beyond the local minimum, with the amount of overshoot tuned to the degree of mesh refinement, yielding more rapid convergence as a result. In the work of R47, Piran and I used SOR in all of our iterative solutions, and achieved substantial gains in convergence speed on our finest meshes.

I became interested in Monte Carlo algorithms because it was clear that lattice gauge theory simulations probably would be the only way that one could study details of the structure of the flux confinement domain in QCD. I knew from talks that I had heard in Princeton that there were two main Monte Carlo methods in use, the Metropolis method and the heat bath method, and also that the folk wisdom at the time was that heat bath was the best one could do, since it corresponded to "nature's way" of achieving thermal equilibrium. However, since the zero temperature limit of

heat bath just corresponds to a Gauss–Seidel iteration, which I knew could be accelerated by SOR, I suspected that the conventional wisdom was wrong, and that there should be extensions, to Monte Carlo thermalizations, of the standard acceleration methods for the solution of differential equations. Since the monopole numerical work had brought out the fact that the Yang–Mills action is multiquadratic, I decided to study this question in the simple context of multiquadratic actions, where the question becomes whether for quadratic actions, the SOR method for differential equations has an extension to Monte Carlo thermalization. This is the question addressed in Adler (1981), R54, where I showed that SOR does indeed extend to the thermalization of multi-quadratic actions, by explicitly constructing in Eqs. (14a,b) the transition probability that obeys detailed balance when an overrelaxation parameter is included in the iteration. (Note that the normalization factors in these equations have the π in the correct place, but the other factors inverted; this is corrected in Eqs. (9) and (11) of my later paper R55. The argument of R54 does not involve the normalization factors, and is unaffected.) For this overrelaxed thermalization, I showed that the means of the thermalized variables iterate according to standard SOR; since standard SOR accelerates Gauss–Seidel, this implies that there should be a corresponding acceleration of the thermalization process as well.

The 1981 paper R54 gave the earliest indication that Monte Carlo methods could be accelerated to improve critical slowing down, and for this reason was conceptually important, as well as having later applications and extensions. To the best of my knowledge, I am supported by the literature on the subject, in stating that R54 first introduced acceleration methods into Monte Carlo. Two compilations of Monte Carlo articles edited by Binder contain literature surveys, Binder et al. (1987), and Swendsen, Wang, and Ferrenberg (1992), relating to the critical slowing down problem. In both surveys, the earliest listed reference is from 1983; neither survey cites my 1981 paper (or Whitmer's 1984 paper – see below), although some of the cited articles do reference these papers.

I didn't immediately continue work on Monte Carlo acceleration myself, but suggested it as a thesis research area to my Princeton University graduate student Charles Whitmer. He applied the method to ϕ^4 and Higgs actions that are point split on a lattice with unit displacement $\hat{\mu}$ according to $\phi^4(x) \to \phi^2(x)\phi^2(x + a\hat{\mu})$, which makes them multiquadratic, and in the paper Whitmer (1984) reported improvement over conventional heat bath Monte Carlo. I got interested in the subject again a few years later, after Goodman and Sokal (1986) (who knew about the SOR method of R54 and Whitmer's paper) proposed a stochastic extension of multigrid methods, and Creutz (1987) and Brown and Woch (1987) gave a simple implementation of the SOR idea for lattice gauge theory plaquette actions. This latter development eliminated the need for the problematic gauge fixing that I had used in R54 to keep the latticized gauge action multiquadratic, and opened the way to practical applications of SOR to gauge theory Monte Carlo studies. In the spring of 1987,

I went to Torino, Italy with my daughter Victoria, who had been eager to visit Europe after finishing her high school requirements. During this sabbatical term I was a visitor at the Institute for Scientific Interchange (ISI), at the invitation of Mario Rasetti and my former IAS colleague Tullio Regge; I also had an office at the University of Torino that I used a couple of days a week. Although I had been spending considerable time over the previous few years working on quaternionic quantum mechanics (see the next chapter), I decided on this trip to return to my old interest in Monte Carlo SOR, stimulated by the fact that experts in the Monte Carlo field had started to get interested. In a paper that I wrote while at ISI (Adler, 1988a, R55) I gave a much more detailed analysis of overrelaxed thermalization for a quadratic action, and also gave extensions of the method to non-quadratic actions, including $SU(n)$ gauge theory.

After my return to the IAS from this sabbatical, I continued work on Monte-Carlo algorithms for several more years. In a paper written in the fall of 1987 after I returned from Italy (Adler, 1988b, R56), I gave an elegant formal analysis showing that the general linear iteration $u' = Mu+Nf$ corresponding to a splitting $1 = M+NL$ of the quadratic form L for a Gaussian action, has a corresponding stochastic generalization

$$P(u \to u') = (\beta/\pi)^{1/2}(\det\Gamma)^{1/2}\exp[-(u' - Mu - Nf)^T\beta\Gamma(u' - Mu - Nf)] \quad ,$$

with $\Gamma = \frac{1}{2}(L^{-1} - ML^{-1}M^T)^{-1}$ a modified temperature matrix. This extends the SOR thermalization of R54 to a general linear iterative process. Later in the same academic year, I gave in Adler (1988c) a Metropolis variant of the $SU(n)$ method given in R55, that extended the method for the Wilson action used by Creutz and by Brown and Woch to general overrelaxation parameter ω. In collaboration with Gyan Bhanot, a former IAS member and a Monte Carlo expert, we made a numerical study of the $SU(2)$ version of this algorithm, with results reported in Adler and Bhanot (1989), R57. (Growing out of this collaboration, Bhanot spent several years as a half-time member of the IAS in the early 1990's, in the course of which we wrote a number of further papers on a variety of Monte Carlo acceleration methods.) I also gave talks at lattice conferences; at the biennial Lattice Gauge conference Lattice 88, held at Fermilab that year, I gave a plenary talk reviewing work on algorithms for pure gauge theory, focusing primarily on the theory and application of overrelaxation methods (Adler, 1989, R58). Monte Carlo overrelaxation has become a standard part of the lattice gauge theorist's tool kit; for a sampling of recent applications, see Kiskis, Narayanan, and Neuberger (2003), Holland, Pepe, and Wiese (2004), Meyer (2004), Pepe (2004), Shcheredin (2005), and de Forcrand and Jahn (2005).

Image Normalization

During the 1990s, I interspersed my work on quaternionic quantum mechanics and particle physics with work on aspects of neural networks and pattern recognition. My neural network interests involved an analog device that I patented (Adler, 1993) and an article (Adler, Bhanot, and Weckel, 1996) analyzing its algorithmic aspects. In pattern recognition, from lunchtime conversations with Joseph Atick and Norman Redlich, I got interested in the problem of extracting those features of an image that are invariant under a symmetry transformation. This problem is closely analogous to that of extracting those features of a gauge potential that are gauge-invariant, and in Adler (1998), R59 I gave a general formal solution, based on imposing image normalizing constraints analogous to gauge-fixing constraints. I have reprinted here only the first two sections of this unpublished article (without references), in which the general theory is set up; further sections of the article give applications to a variety of viewing transformations of a planar object. Shortly afterwards, when one of the IAS string theory postdocs was interested in switching to a computer-related career, I suggested applying my methods to the problem of the similarity and affine normalization of partially occluded planar curves (such as the boundary of a planar object). We worked this out together and it was published as Adler and Krishnan (1998), R60. The excerpt R59 of the general paper that is reprinted here gives the background needed to follow the extension of the planar algorithm to curve segments given in R60.

References for Chapter 8

Adler, S. L. (1981) R54. Over-Relaxation Method for the Monte Carlo Evaluation of the Partition Function for Multiquadratic Actions. *Phys. Rev. D* **23**, 2901-2904. [99,100]

Adler, S. L. (1988a) R55. Overrelaxation Algorithms for Lattice Field Theories. *Phys Rev. D* **37**, 458-471. [99,100]

Adler, S. L. (1988b) R56. Stochastic Algorithm Corresponding to a General Linear Iterative Process. *Phys. Rev. Lett.* **60**, 1243-1245. [100]

Adler, S. L. (1988c). Metropolis Overrelaxation for Lattice Gauge Theory for a General Relaxation Parameter ω. *Phys. Rev. D* **38**, 1349-1351. [100]

Adler, S. L. (1989) R58. Algorithms for Pure Gauge Theory. *Nucl. Phys. B Proc. Suppl.* **9**, 437-446. [100]

Adler, S. L. (1993). Neural Network Architecture Based on Summation of Phase-Coherent Alternating Current Signals. U.S. Patent No. 5,261,035. [101]

Adler, S. L. (1998) R59. General Theory of Image Normalization, unpublished; arXiv: cs.CV/9810017. A slightly abridged and reformatted version of the first two sections is reprinted here. [101]

Adler, S. L. and G. V. Bhanot (1989) R57. Study of an Overrelaxation Method for Gauge Theories. *Phys. Rev. Lett.* **62**, 121-124. [100]

Adler, S. L., G. Bhanot, and J. D. Weckel (1996). Algorithmic Aspects of a Neuron for Coherent Wave Synapse Realizations. *IEEE Trans. Neural Networks* **7**, 1262-1271. [101]

Adler, S. L. and R. Krishnan (1998) R60. Similarity and Affine Normalization of Partially Occluded Planar Curves Using First and Second Derivatives. *Pattern Recognition* **31**, 1551-1556. [101]

Binder, K. et al. (1987). Recent Developments, in *Applications of the Monte Carlo Method in Statistical Physics*, K. Binder, ed. (Springer-Verlag, Berlin), Chapter 10, p. 302, refs. [10.15-22]. [99]

Brown, F. R. and T. J. Woch (1987). Overrelaxed Heat-Bath and Metropolis Algorithms for Accelerating Pure Gauge Monte Carlo Calculations. *Phys. Rev. Lett.* **58**, 2394-2396. [99,100]

Creutz, M. (1987). Overrelaxation and Monte Carlo Simulation. *Phys. Rev. D* **36**, 515-519. [99,100]

de Forcrand, P. and O. Jahn (2005). Monte Carlo Overrelaxation for $SU(N)$ Gauge Theories; arXiv: hep-lat/0503041. [100]

Goodman, J. and A. D. Sokal (1986). Multigrid Monte Carlo Method for Lattice Field Theories. *Phys. Rev. Lett.* **56**, 1015-1018. [99]

Holland, K., M. Pepe, and U. J. Wiese (2004). The Deconfinement Phase Transition of $SP(2)$ and $SP(3)$ Yang-Mills Theories in (2+1)-Dimensions and (3+1)-Dimensions. *Nucl. Phys. B* **694**, 35-58. [100]

Kiskis, J., R. Narayanan, and H. Neuberger (2003). Does the Crossover from Perturbative to Nonperturbative Physics in QCD Become a Phase Transition at Infinite N? *Phys. Lett. B* **574**, 65-74. [100]

Meyer, H. B. (2004). The Yang-Mills Spectrum from a Two-Level Algorithm. *JHEP* **0401**, 030. [100]

Pepe, M. (2004). Deconfinement in Yang-Mills: A Conjecture for a General Gauge Lie Group G. *Nucl. Phys. B. Proc. Suppl.* **141**, 238-243 (2005). [100]

Shcheredin, S. (2005). Simulations of Lattice Fermions with Chiral Symmetry in Quantum Chromodynamics; arXiv: hep-lat/0502001. [100]

Swendsen, R. H., J.-S. Wang, and A. M. Ferrenberg (1992). New Monte Carlo Methods for Improved Efficiency of Computer Simulations in Statistical Physics, in *The Monte Carlo Method in Condensed Matter Physics*, K. Binder, ed. (Springer-Verlag, Berlin), Sec. 4.1, pp. 75-76. [99]

Whitmer, C. (1984). Over-Relaxation Methods for Monte Carlo Simulations of Quadratic and Multiquadratic Actions. *Phys. Rev. D* **29**, 306-311. [99]

9. Quaternionic Quantum Mechanics, Trace Dynamics, and Emergent Quantum Theory

Introduction

During the twenty years from 1984 to 2004, a large part of my time was spent on investigations into foundational areas of quantum mechanics. Most of my research from this period was later presented in two books that I wrote, *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford University Press, New York, 1995) and *Quantum Theory as an Emergent Phenomenon* (Cambridge University Press, Cambridge, 2004). I have not included in this reprint volume any research papers incorporated (some considerably improved) into the two books, since this would be infeasible because of length limitations. So what I discuss in this chapter are a few papers dealing with quaternionic topics written during the period between the two books, together with a brief description of how I got interested in quaternionic quantum theory, and later on, in the possibility of a pre-quantum theory.

Quaternionic Quantum Mechanics

My interest in quaternionic quantum mechanics grew out of my interest in the Harari (1979)-Shupe (1979) model for composite quarks and leptons. They postulated an order-dependence for the preon wave functions (e.g., TTV, TVT, VTT were considered to be three distinct color states), which suggested that quantum theory over a noncommutative field might be involved. I was never able to use quaternions or related ideas to make a successful preon model, either during the period before my book (see Adler, 1979, 1980) or after (Adler, 1994a), but the issues raised, and interactions with key people acknowledged in the Preface of the 1995 volume, led me to undertake a systematic study of quaternionic quantum mechanics. Perhaps the most important new result contained in my papers (Adler, 1988) and in my book is the fact that the S-matrix in quaternionic scattering theory is *complex*, not quaternionic, which was a surprise to the experts in the field and invalidated proposed searches (such as Peres, 1979) for quaternionic effects manifested through noncommuting scattering phases. I also clarified the relationship between time reversal symmetry in quaternionic quantum theory (where it is unitary) and in complex quantum theory (where it is antiunitary), proved that positive energy quaternionic Poincaré group representations are complex and not intrinsically quaternionic, and

gave a quaternionic generalization of projective group representations (to which I shall return shortly). These were but a few of the many topics dealt with in my 1995 book. My quaternionic investigations also motivated work I did in new directions in standard quantum mechanics, such as a paper that I wrote showing that $SU(3) \times SU(12)$ is the minimal grand unified theory in which, species by species for charged fermions, no Dirac sea is required (Adler, 1989).

After my book on quaternionic quantum mechanics was completed, a number of papers that I wrote with collaborators clarified issues that were left unresolved, or were inadequately treated, in the book. One of these issues dealt with the non-adiabatic geometric phase in quaternionic Hilbert space. This was discussed in my book, but on a visit to the IAS, Jeeva Anandan pointed out that my treatment was incomplete, and sketched what was needed to improve it. I filled in the details and drafted a manuscript, which became a joint paper (Adler and Anandan, 1996, R61) that was published in the Larry Horwitz Festschrift issue of *Foundations of Physics*. A second issue that was left hanging was the analog of coherent states in quaternionic quantum theory. My thesis student Andrew Millard and I studied this, and wrote a paper (Adler and Millard, 1996a, R62) giving the extension of the Perelomov coherent state formalism to quaternionic Hilbert spaces. We also showed that the closure requirement forces an attempted quaternionic generalization of standard coherent states based on the Weyl group to reduce back to the complex case, settling a question raised in discussions with me by John Klauder. The other issues that were dealt with after publication of the quaternionic book were the structure of quaternionic projective representations, and the relationship between standard complex quantum mechanics and the dynamics based on a trace variational principle that I had introduced in the field theory chapter of the 1995 book. These form the subject of the next two sections.

Quaternionic Projective Group Representations

Given two group elements b, a with product ba, a unitary operator representation U_b in a Hilbert space is defined by $U_b U_a = U_{ba}$. A more general type of representation, called a ray or projective representation, is relevant to describing the symmetries of quantum mechanical systems. In his famous paper on unitary ray representations of Lie groups, Bargmann (1954) defines a projective representation as one obeying $U_b U_a = U_{ba}\omega(b,a)$, with $\omega(b,a)$ a complex phase.

This definition is familiar, and seems obvious, until one asks the following question: Bargmann's definition is assumed to hold as an operator identity when acting on *all* states in Hilbert space. However, we know that it suffices to specify the action of an operator on *one* complete set of states in Hilbert space to specify the operator completely. Hence why does one not start instead from the definition $U_b U_a|f\rangle = U_{ba}|f\rangle\omega(f;b,a)$, with $\{|f\rangle\}$ one complete set of states, as defining a pro-

jective representation in Hilbert space? Let us call Bargmann's definition a "strong" projective representation, and the definition with a state-dependent phase a "weak" projective representation. Then the question becomes that of finding the relation between weak and strong projective representations.

Although I have formulated this question here in complex Hilbert space, it arose and was solved in the context of quaternionic Hilbert space, where the phases $\omega(f;b,a)$ are quaternions, which obey a non-Abelian group multiplication law isomorphic to $SO(3) \simeq SU(2)$. The strong definition was adopted for the quaternionic case by Emch (1963, 1965), but in Sec. 4.3 of my book on quaternionic quantum mechanics I introduced the weak definition in order for quaternionic projective representations to include embeddings of nontrivial complex projective representations into quaternionic Hilbert space; the state dependence of the phase is necessary because even a complex phase ω does not commute with general quaternionic rephasings of the state vector $|f\rangle$. I noted in my 1995 book that the weak definition can be extended to an operator relation by defining

$$\Omega(b,a) = \sum_f |f\rangle\omega(f;b,a)\langle f| \quad ,$$

so that the weak definition then takes the form

$$U_b U_a = U_{ba}\Omega(b,a) \quad ,$$

which gives the general operator form taken by projective representations in quaternionic quantum mechanics. I also introduced in Sec. 4.3 of my book two specializations of this definition, motivated by the commutativity properties of the phase factor in complex projective representations. I defined a *multicentral* projective representation as one for which

$$[\Omega(b,a), U_a] = [\Omega(b,a), U_b] = 0$$

for all pairs b, a (note that in Eq. (4.51a) of my book, U_{ab} should read U_{ba}, so that the two conditions just given suffice), and I defined a *central* projective representation as one for which

$$[\Omega(b,a), U_c] = 0$$

for all triples a, b, c.

Subsequent to the completion of my book, I read Weinberg's first volume on quantum field theory (Weinberg, 1995) and realized, from his discussion in Sec. 2.7 of the associativity condition for complex projective representations, that there must be an analogous associativity condition for weak quaternionic projective representations. I worked this out (Adler, 1996, R63), and showed that it takes the operator form

$$U_a^{-1}\Omega(c,b)U_a = \Omega(cb,a)^{-1}\Omega(c,ba)\Omega(b,a) \quad ,$$

which by the definition of $\Omega(b,a)$ shows that $U_a^{-1}\Omega(c,b)U_a$ is diagonal in the basis $\{|f\rangle\}$, with the spectral representation

$$U_a^{-1}\Omega(c,b)U_a = \sum_f |f\rangle\overline{\omega(f;cb,a)}\omega(f;c,ba)\omega(f;b,a)\langle f| \quad .$$

On the basis of some further identities, I also raised the question of whether one can construct a multicentral representation that is not central, or whether a multicentral representation is always central.

Subsequently, I discussed the issues of quaternionic projective representations with Andrew Millard. He explained them to his roommate Terry Tao, a mathematics graduate student working for Elias Stein, and at my next conference with Andrew, Tao came along and presented the outline of a beautiful theorem that he had devised. This was written up as a paper of Tao and Millard (1996), and consists of two parts. The first part is an algebraic analysis based on the spectral representation given above, which leads to the following theorem

Structure Theorem: *Let U be an irreducible projective representation of a connected Lie group G. There then exists a reraying of the basis $|f\rangle$ under which one of the following three possibilities must hold.*

(1) U is a real projective representation. That is, $\omega(f;b,a) = \omega(b,a)$ is independent of $|f\rangle$ and is equal to ± 1 for each b and a.

(2) U is the extension of a complex projective representation. That is, the matrix elements $\langle f|U_a|f'\rangle$ are complex and $\omega(f;b,a) = \omega(b,a)$ is independent of $|f\rangle$ and is a complex phase.

(3) U is the tensor product of a real projective representation and a quaternionic phase. That is, there exists a decomposition $U_a = U_a^{\mathcal{B}}\sum_f |f\rangle\sigma_a\langle f|$, where the unitary operator $U_a^{\mathcal{B}}$ has real matrix elements, σ_a is a quaternionic phase, and $U_{ba}^{\mathcal{B}} = \pm U_b^{\mathcal{B}}U_a^{\mathcal{B}}$ for all b and a.

From the point of view of the Structure Theorem, case (1) corresponds to the only possibility allowed by the strong definition of quaternionic projective representations, as demonstrated earlier by Emch (1963, 1965), while case (2) corresponds to an embedding of a complex projective representation in quaternionic Hilbert space, the consideration of which was my motivation for proposing the weak definition. Specializing the Structure Theorem to a complex Hilbert space, where case (3) cannot be realized, we see that in complex Hilbert space the weak projective representation defined above *implies* the strong projective representation; hence no generality is lost by starting from the strong definition, as in Bargmann's paper.

The second part of the Tao–Millard paper is a proof, by real analysis methods, of a Corollary to the structure theorem, stating

Corollary 1: *Any multicentral projective representation of a connected Lie group is central.*

This thus solved the question of the relation of centrality to multicentrality that I raised in my paper R63.

Subsequent to this work, I had an exchange with Gerard Emch in the *Journal of Mathematical Physics* debating the merits of the strong and weak definitions. After a visit to Gainesville where we reconciled differing notations, we wrote a joint paper (Adler and Emch, 1997, R64) clarifying the situation, and reexpressing the strong and weak definitions in the language and notation often employed in mathematical discussions of projective group representations.

Trace Dynamics and Emergent Quantum Theory

My work on emergent quantum theory arose from the merging of two lines of thought. The first line of thought arose from answering the question of whether quaternionic quantum mechanics ameliorates the measurement problem of standard quantum mechanics; the answer is "no", because quaternionic quantum theory still has a unitary time evolution, and so the usual problems persist. However, in the course of working this through I read some of the literature on the measurement problem in standard quantum theory, and came away convinced that there were real issues to be addressed. The second line of thought arose from my attempts to construct quaternionic quantum field theories. I found that the canonical quantization method could not be extended to the quaternionic case, and so I had to resort to an alternative formalism, which I variously called "generalized quantum dynamics", "total trace dynamics", or finally, simply "trace dynamics", to generate operator equations without "quantizing" a classical theory. This was done by using a variational principle based on a Lagrangian constructed as a trace of noncommuting operator variables, making systematic use of cyclic permutation under the trace operation. These ideas were developed in the paper Adler (1994b) and were described in Chapter 13 of my 1995 book; in Chapter 14, I suggested that the nonlinearity of trace dynamics could make it relevant for resolving the measurement problem in quantum theory. However, the problem of relating the trace dynamics formalism to the standard canonical formalism of complex quantum field theory remained unsolved.

One of the questions I had posed to Andrew Millard was that of better understanding trace dynamics, in the hope of finding a connection to standard quantum theory. After I arrived in Aspen in the summer of 1995, Andrew sent me a memo containing his discovery that in trace dynamics with a Weyl symmetrized Hamiltonian and noncommuting boson degrees of freedom q_r, p_r, the operator $\sum_r [q_r, p_r]$ is conserved. I soon found that the generalization to include fermions is the conserved

operator that we denoted by $\tilde{C}$, defined by

$$\tilde{C} \equiv \sum_{r,\,\text{boson}} [q_r, p_r] - \sum_{r,\,\text{fermion}} \{q_r, p_r\} \quad ,$$

and that this operator is conserved as long as the trace Hamiltonian has no fixed operator coefficients, which is equivalent to saying the the trace Hamiltonian has a global unitary invariance. It then seemed natural to suggest that the equipartitioning of $\tilde{C}$ in a statistical thermodynamical treatment would provide the missing connection between trace dynamics and standard quantum mechanics.

The implementation of this idea was published in Adler and Millard (1996b), and I developed it further over the following years with many collaborators, as described in Sec. 5 of the "Introduction and Overview" that opens my 2004 book on emergent quantum theory. This book, which is set within the framework of complex Hilbert space, gives a complete, self-contained development of trace dynamics as a (noncommutative) dynamics underlying quantum theory. From the statistical mechanics of this underlying theory there emerge, in a mutually complementary way, both the unitary and the nonunitary parts of orthodox quantum theory. The unitary part of quantum theory (the canonical algebra and the Heisenberg representation time evolution of operators) comes from an application of generalized equipartition theorems in the statistical thermodynamics of trace dynamics. The nonunitary part of quantum theory, in the form of stochastic state vector reduction models from which the Born rule for probabilities can be derived, comes from the Brownian motion corrections to this thermodynamics. Thus, trace dynamics provides a unified framework from which both the unitary dynamics of quantum systems, and the nonunitary evolution describing state vector reduction associated with measurements, emerge in a natural way.

Although quantum mechanics and quantum field theory have been the undisputed basis for all progress in fundamental physics during the last 80 years, the extension of the current theoretical frontier to Planck scale physics, and recent enlargements of our experimental capabilities, may make the 21st century the period in which possible limits of quantum theory will be probed. My 2004 book suggests a concrete framework for exploration of the proposition that quantum mechanics may not be the final layer of fundamental theory. It also addresses the phenomenology of modifications to quantum theory, specifically as implemented through stochastic additions to the Schrödinger equation. I have continued with these phenomenological studies since completion of the book; my most recent papers (Bassi, Ippoliti, and Adler, 2005; Adler, Bassi, and Ippoliti, 2005; Adler, 2005) have dealt with analyzing possible tests of stochastic localization theories in nanomechanical oscillator and gravitational wave detector experiments.

References for Chapter 9

Adler, S. L. (1979). Algebraic Chromodynamics. *Phys. Lett. B* **86**, 203-205. [103]

Adler, S. L. (1980). Quaternionic Chromodynamics as a Theory of Composite Quarks and Leptons. *Phys. Rev. D* **21**, 2903-2915. [103]

Adler, S. L. (1988). Scattering and Decay Theory for Quaternionic Quantum Mechanics, and the Structure of Induced T Nonconservation. *Phys. Rev. D* **37**, 3654-3662. [103]

Adler, S. L. (1989). A New Electroweak and Strong Interaction Unification Scheme. *Phys. Lett. B* **225**, 143-147. [104]

Adler, S. L. (1994a). Composite Leptons and Quarks Constructed as Triply Occupied Quasiparticles in Quaternionic Quantum Mechanics. *Phys. Lett. B* **332**, 358-365. [103]

Adler, S. L. (1994b). Generalized Quantum Dynamics. *Nucl. Phys. B* **415**, 195-242. [107]

Adler, S. L. (1996) R63. Projective Group Representations in Quaternionic Hilbert Space *J. Math. Phys.* **37**, 2352-2360. [105-107]

Adler, S. L. (2005). Stochastic Collapse and Decoherence of a Non-Dissipative Forced Harmonic Oscillator. *J. Phys. A: Math. Gen.* **38**, 2729-2745. [108]

Adler, S. L. and J. Anandan (1996) R61. Nonadiabatic Geometric Phase in Quaternionic Hilbert Space. *Found. Phys.* **26**, 1579-1589. [104]

Adler, S. L., A. Bassi, and E. Ippoliti (2005). Towards Quantum Superpositions of a Mirror: An Exact Open Systems Analysis - Calculational Details. *J. Phys. A: Math. Gen.* **38**, 2715-2727. [108]

Adler, S. L. and G. G. Emch (1997) R64. A Rejoinder on Quaternionic Projective Representations. *J. Math. Phys.* **38**, 4758-4762. [107]

Adler, S. L. and A. C. Millard (1996a) R62. Coherent States in Quaternionic Quantum Mechanics. *J. Math. Phys.* **38**, 2117-2126. [104]

Adler, S. L. and A. C. Millard (1996b). Generalized Quantum Dynamics as Pre-Quantum Mechanics. *Nucl. Phys B* **473**, 199-244. [108]

Bargmann, V. (1954). On Unitary Ray Representations of Continuous Groups. *Ann. Math.* **59**, 1-46. [104-106]

Bassi, A., E. Ippoliti, and S. L. Adler (2005). Towards Quantum Superpositions of a Mirror: An Exact Open Systems Analysis. *Phys. Rev. Lett.* **94**, 030401. [108]

Emch, G. (1963). Mécanique Quantique Quaternionienne et Relativité Restreinte. I, *Helv. Phys. Acta* **36**, 739-769; II, *Helv. Phys. Acta* **36**, 770-788. [105,106]

Emch, G. (1965). Representations of the Lorentz Group in Quaternionic Quantum Mechanics (presented at the summer 1964 Lorentz Group Symposium), in *Lectures in Theoretical Physics* Vol. VIIA, W.E. Brittin and A. O. Barut, eds. (University of Colorado Press, Boulder), pp. 1-36. [105,106]

Harari, H. (1979). A Schematic Model of Quarks and Leptons. *Phys. Lett. B* **86**, 83-86. [103]

Peres, A. (1979). Proposed Test for Complex versus Quaternion Quantum Theory. *Phys. Rev. Lett.* **42**, 683-686. [103]

Shupe, M. A. (1979) A Composite Model of Leptons and Quarks. *Phys. Lett. B* **86**, 87-92. [103]

Tao, T. and A. C. Millard (1996). On the Structure of Projective Group Representations in Quaternionic Hilbert Space. *J. Math. Phys.* **37**, 5848-5857. [106]

Weinberg, S. (1995) *The Quantum Theory of Fields*, Volume I: Foundations (Cambridge University Press, Cambridge), Sec. 2.7. [105]

10. Where Next?

In looking back at my work, I see one pattern that is repeated over and over. Many of the most interesting research results that I have obtained were unanticipated consequences of other, quite different research programs. In the course of detailed calculations, or speculative explorations, I noticed something that seemed worth pursuing, even though tangential to my original motivations, and this new direction ended up being of much greater interest. This happened with my calculations of weak pion production, which led as spin-offs to the forward lepton theorem, the neutrino sum rule, and soft pion theorems. It happened again with my exploration of gauging of the axial-vector current as an explanation for the muon mass, which led to anomalies. My interest in an eigenvalue condition in QED led to the calculation of photon splitting, and later on to an improved method for analyzing collider data. My attempts at a composite graviton led to an investigation of Einstein gravity as a symmetry breaking effect. My interest in the (spurious) Argonne threshold events induced me to extend my weak pion work to neutral currents, which contributed to the first unique determination of the electroweak couplings by Abbott and Barnett. My attempt to relate monopole background fields to confinement played a role in the multimonopole existence proof of Taubes. My computational experience in solving effective action confinement models led to overrelaxation as an acceleration method for Monte Carlo. And most recently, my interest in composite models for quarks and leptons led to a long exploration of the fundamentals of quantum theory, first through my study of quantum theory in quaternionic Hilbert space, and growing out of that, through my development of trace dynamics as a possible pre-quantum theory.

I think this pattern is no accident, but rather a reflection of my guiding philosophy in doing research, which has been that it is more important to start somewhere, even with a speculative idea or an apparently routine calculation, than to sit around waiting for an "important" idea. Once immersed in the nitty-gritty of an investigation, things have a way of appearing, that often lead off in very fruitful directions. So given this, when I look ahead, I can only say the following: I have some rough ideas as to where I would like to start in new explorations in fundamental theory and particle phenomenology, but I cannot say where these may ultimately lead, in the course of my continuing adventures in theoretical physics.

Index of Names in the Commentaries

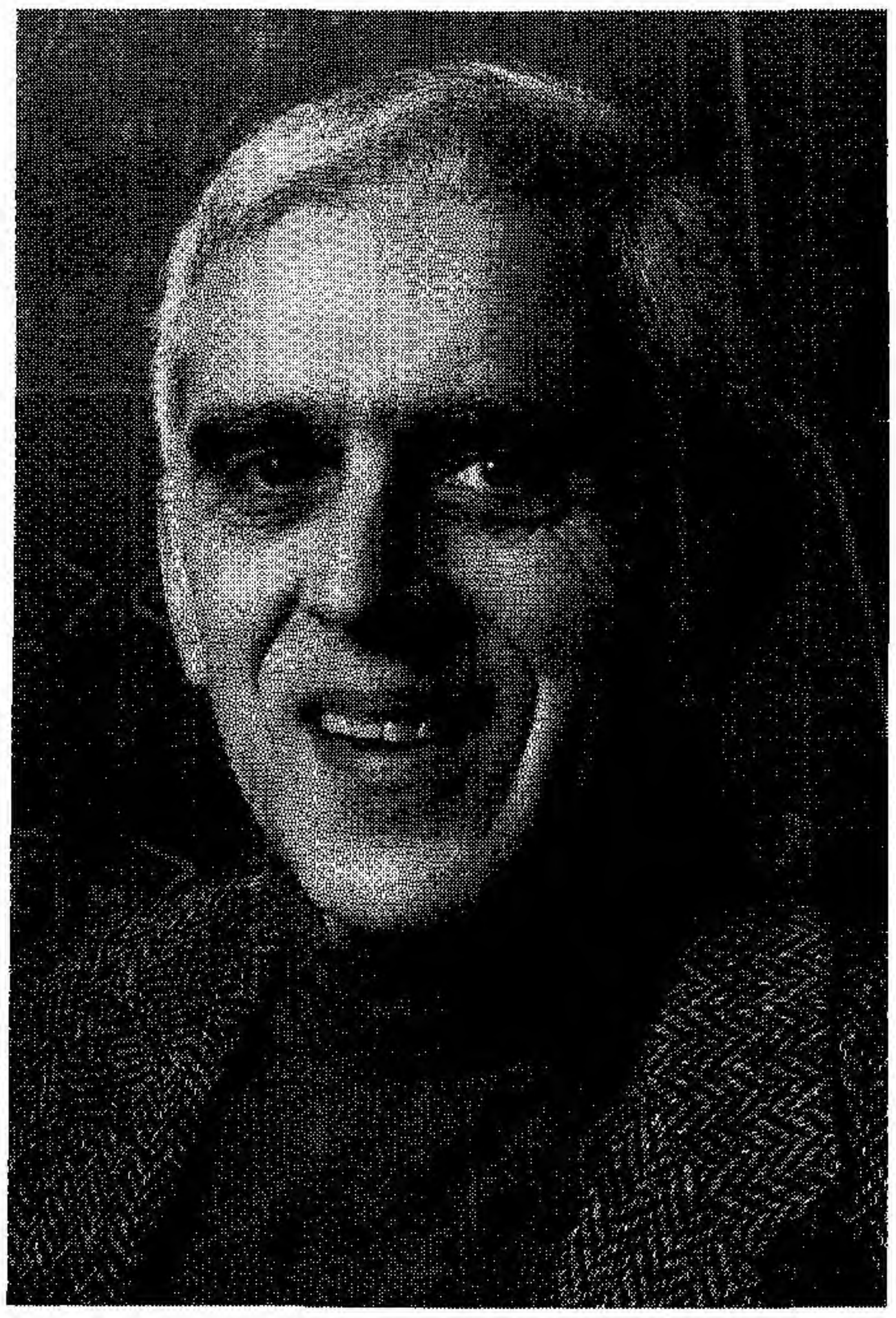
This page is intentionally left blank

REPRINTED PAPERS

This page is intentionally left blank

FROM ELEMENTS OF RADIO TO ELEMENTARY PARTICLE PHYSICS

Stephen L. Adler
Institute for Advanced Study at Princeton, USA



© Courtesy of the Institute for Advanced Study

I was born in 1939 in New York City to Irving and Ruth Relis Adler. My father was a mathematics teacher and my mother had also majored in mathematics in college. I was directed towards science by my parents from an early age. When I was two years old my father built me a gadget box from pieces of hardware, and around the same time my mother made me a home-made version of the "Pat the Bunny" book, each page containing a different tactile or manual operation for me to perform. When I was older my father built me electrical toys such as telegraphs, a "burglar alarm" that rang a bell when a door was opened, and a miniature traffic light. We also engaged in nature activities, collecting snakes and butterflies. When I was eight I participated in a young people's astronomy course at the Museum of Natural History in New York, and my fascination with the fossils I saw at this museum led me to think briefly of being a paleontologist, but this interest soon waned.

My actual career path began in sixth grade of elementary school, when a classmate started to talk to me about his interest in radio; I visited him at home and saw his equipment and tools. This introduction developed into a serious interest in electricity, radio, and electronics while I was still in elementary school. I built various electrical devices, such as electric motors with rotor laminations cut from tin cans and permanent magnet stators taken from old radio loudspeakers. (I still have one of these on my bookcase at the Institute for Advanced Study). With encouragement from my father I read Marcus and Marcus's classic World War II text "Elements of Radio"; my father made a point of letting me be the family radio expert, while he was the consultant on the few bits of algebra in the text. Also at my father's suggestion, I started to canvass the neighborhood door to door, pulling a small wagon and asking for old radios, appliances, and television sets people were planning to throw away. I stripped the parts out of these, and used them to build radios, amplifiers, and even an oscilloscope using a salvaged 7-inch television tube. I also learned enough Morse code to get my Technician Class amateur radio license, and built a small rig using a war surplus aircraft receiver and a home-built transmitter. However, amateur radio activity did not interest me nearly as much as building electronic equipment, which I continued through various science projects in high school.

Given this exposure to electronics, it would have been natural for me to pursue a career in electrical engineering, but towards the beginning of my high school years I got a first glimpse of the fascinating world of high energy physics research. For two summers my family had vacationed in a state park near Ithaca, NY, and Phillip Morrison, an old friend of my father's, gave us a tour of the physics laboratories at Cornell, where Robert Wilson had built a succession of particle accelerators. I liked the ambience of these laboratories, and was impressed with the fact that if I pursued physics as a career I would learn and use electronics, but not necessarily the other way around. By my junior year in high school, I had decided on experimental physics as a career.

My first physics research laboratory experience came at the end of my senior year in high school, when I attended a two-week course in X-ray diffraction techniques for industrial engineers given at Brooklyn Polytechnic Institute by Isadore Fankuchen, who would every now and then include a bright high school student in his class. I was able to do all the theoretical and laboratory work, and learned many things, such as crystal lattice structure and Fourier transforms, that are standard physicist's tools. Almost immediately afterwards, I went to a summer job at Bell Labs in Manhattan, along with eight other science-oriented high

school graduates. Many of them had already learned calculus, and so I decided to teach myself calculus that summer.

My father gave me his old calculus text, along with the sage advice to do every *third* problem—because I had to do problems to learn the material, but there was not time (and it would be too boring) to try to do all of them. So I spent my commuting time, and spare time at work, doing calculus problems. As a result, when I entered Harvard in the fall I was able to place directly into Advanced Calculus, which had a major impact in how fast I was able to proceed with my physics education.

I entered college intending to be an experimentalist, but my friendships with various classmates, among them Daniel Quillen (later a Fields medalist) got me interested in mathematics. I found that I was very good at the theoretical aspects of my classes, but although competent in the laboratory, I lacked the touch of the gifted experimentalist. So, by the middle of my freshman year, I had decided to shift my sights from experimental to theoretical physics. Along with Fred Goldhaber, who was to be my first year roommate in graduate school at Princeton, I took essentially the whole graduate course curriculum at Harvard during my junior and senior years. Memorable teachers at Harvard included Ed Purcell, Frank Pipkin, Paul Martin, and Julian Schwinger. As a result of my Harvard preparation, at Princeton I was able to take my General Exams at the end of the first year, and then to start thesis research with Sam Treiman at the beginning of my second year.

Treiman suggested that I look for calculations to do in the newly emerging area of accelerator neutrino experiments, and this was the beginning of my career in high energy physics. A major part of my thesis work was a calculation of pion production from nucleons (protons or neutrons) by a neutrino beam. Although this was a long and tedious project, it gave me a detailed introduction to the "vector" and "axial-vector" currents through which neutrinos interact with nucleons. This knowledge growing out of my thesis project was the foundation for my most significant scientific contributions during the period 1964 through 1972, which all involved in some way or another the discovery of further results connected with vector and axial-vector currents. These included various low energy theorems for pion emission derived from the hypothesis of a "partially conserved" axial vector current, various sum rules including the Adler-Weisberger sum rule for the axial vector coupling to nucleons and a sum rule for deep inelastic high energy neutrino scattering cross sections, as well as the co-discovery (along with Bell and Jackiw) of the "anomalous" divergence properties of the axial-vector current. The theoretical analysis of anomalies led to a deeper understanding of neutral pion decay into gamma rays, provided one of the first pieces of evidence for the fact that each quark comes in three varieties (now called "colors"), and has had a multitude of other consequences for theoretical physics over the last thirty-five years.

During the years since 1972, I have worked on a variety of other topics in theoretical high energy physics, including neutral current phenomenology, strong field electromagnetic processes (such as photon splitting near pulsars), and acceleration methods for Monte Carlo simulation algorithms. Throughout the last twenty years I have devoted about half of my research time to studying embeddings of standard quantum mechanics in larger mathematical frameworks. One aspect of this work involved a detailed study of a quantum mechanics in which quaternions replace the usual complex numbers. Another, more recent aspect, has involved the study a possible "pre-quantum" mechanics based on properties of the trace of a matrix, from which quantum theory can emerge as a form of thermodynamics. I have written books describing both of these studies. For the next few years, I plan to return to my original area of particle phenomenology, in the context of supersymmetric models for further unifying the elementary particles and the forces acting on them.

PHYSICAL REVIEW VOLUME 126, NUMBER 1 APRIL 1, 1962

Theory of the Valence Band Splittings at k=0 in Zinc-Blende and Wurtzite Structures

STEPHEN L. ADLER*
General Telephone & Electronics Laboratories, Inc., Bayside Laboratories, Bayside, New York
(Received September 28, 1961)

A theory of the valence band splittings at $\mathbf{k}=0$ in zinc-blende and wurtzite structures is proposed, in which the wurtzite levels are treated as perturbations of those in zinc blende. Starting from one-electron Hamiltonians for the two structures, the two-parameter formulas originally derived by Hopfield are obtained, with a minimum of approximations, along with explicit expressions for the parameters in terms of Hamiltonian matrix elements. The two-parameter formulas are compared with experimental data and agreement is found to be good. A simple tight-binding (linear combination of atomic orbitals), $3p$ valence band, point-ion lattice model is used to calculate an effective charge for ZnS from the known valence band splittings in the wurtzite and zinc-blende dimorphs; a value of $2.3e$ is obtained.

1. INTRODUCTION

THE object of this paper is to explore a theory of the $\mathbf{k}=0$ valence band energy splittings and wave functions in zinc-blende and wurtzite structures. This is of interest because recent experimental work on hexagonal (wurtzite) and cubic (zinc blende) ZnS,[1] hexagonal CdS,[1] CdSe,[2] and ZnO,[3] cubic ZnSe,[4] and other II–VI semiconductors has made available data on their valence band energy splittings and wave function symmetries.

Previous theoretical work on wurtzite and zinc-blende valence band splittings has been reported by Birman[5] and by Hopfield.[6] Birman's theory did not include the spin-orbit interaction. Hopfield's work, based on a quasi-cubic model of the wurtzite structure, gave useful formulas for the wurtzite and zinc-blende valence band splittings, which fit experimental data obtained for ZnS to within 10%. The theory presented below starts from the rigorous one-electron Hamiltonians for wurtzite and zinc blende and yields, with a minimum of approximations, the formulas proposed by Hopfield. In addition, the crystal field and spin-orbit parameters are expressed in terms of matrix elements of the wurtzite and zinc-blende Hamiltonians.

Although the anion and cation are for the sake of definiteness assumed to be S and Zn, this does not enter into the derivation. The two-parameter formulas should be valid in other substances than ZnS crystallizing in zinc-blende and wurtzite dimorphs, provided that the approximations made in the derivation still hold.

* Summer visitor, 1960; now at Physics Department, Princeton University, Princeton, New Jersey.
[1] J. L. Birman, H. Samelson, and A. Lempicki, GT&E R&D J. 1, 1 (1961).
[2] J. O. Dimmock and R. G. Wheeler, J. Appl. Phys. 32, 2271S (1961).
[3] D. G. Thomas, J. Phys. Chem. Solids 15, 86 (1960).
[4] M. Aven, D. T. F. Marple, and B. Segall, General Electric Research Laboratory Report 61-RL-(2773G) (unpublished).
[5] J. L. Birman, Phys. Rev. 115, 1493 (1959).
[6] J. J. Hopfield, J. Phys. Chem. Solids 15, 97 (1960).

2. THEORY

Some details of the zinc-blende and wurtzite geometries will be needed.[7] In zinc blende there are two atoms per unit cell, and in wurtzite, four. Call the two sulfur atoms in the wurtzite basis sulfur 1 and 2, respectively. The nearest-neighbor configurations in zinc blende and ideal wurtzite are identical and the second-nearest-neighbor (nearest like ion) configurations are nearly so (Fig. 1). This local structural similarity will be exploited by using axis systems for the two structures in which the nearest neighbors of the sulfur 1 site in the wurtzite basis and the sulfur site in the zinc blende basis have identical coordinates. The axis systems which will be used throughout the rest of this paper, unless explicitly stated otherwise, are illustrated in Fig. 2. Note that the axes for zinc blende are not the usual Cartesian axes employed for this structure.[5]

The Hamiltonians for band-theoretic treatment of wurtzite and zinc blende, including in each the spin-orbit interaction term, are

$$H_{\mathrm{ZB}}=\mathbf{p}^2/2m+V_{\mathrm{ZB}}(\mathbf{r})+(\hbar/4m^2c^2)(\nabla V_{\mathrm{ZB}}\times\mathbf{p})\cdot\boldsymbol{\sigma},$$
$$H_{\mathrm{W}}=\mathbf{p}^2/2m+V_{\mathrm{W}}(\mathbf{r})+(\hbar/4m^2c^2)(\nabla V_{\mathrm{W}}\times\mathbf{p})\cdot\boldsymbol{\sigma},$$

where V_{ZB} and V_{W} are the respective crystal potentials in the two substances. A modification of the linear combination of atomic orbitals (LCAO) procedure will be used to find expressions for the valence band energy splittings and wave functions at $\mathbf{k}=0$ in the Brillouin zone. In the usual LCAO formalism,[8] the Hamiltonian at $\mathbf{k}=0$ is diagonalized in a space spanned by the cell periodic functions $\Psi_n{}^\beta=N^{-\frac{1}{2}}\sum_j\psi_n(\mathbf{r}-\mathbf{R}_j{}^\beta)$. Here ψ_n is a free-ion orbital, n denotes a complete set of quantum numbers and $\mathbf{R}_j{}^\beta$ is the position of the βth basis atom in the jth unit cell. In the modification used in this paper, the rigorous zinc-blende (ZB) Bloch functions at $\mathbf{k}=0$ are expanded in Wannier functions,

$$\Psi_n{}^{\mathrm{ZB}}=N^{-\frac{1}{2}}\sum_j\chi_n(\mathbf{r}-\mathbf{R}_j{}^1),$$

where $\mathbf{R}_j{}^1$ is the coordinate of the sulfur atom in the

[7] J. L. Birman, Phys. Rev. 109, 810 (1958).
[8] J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).

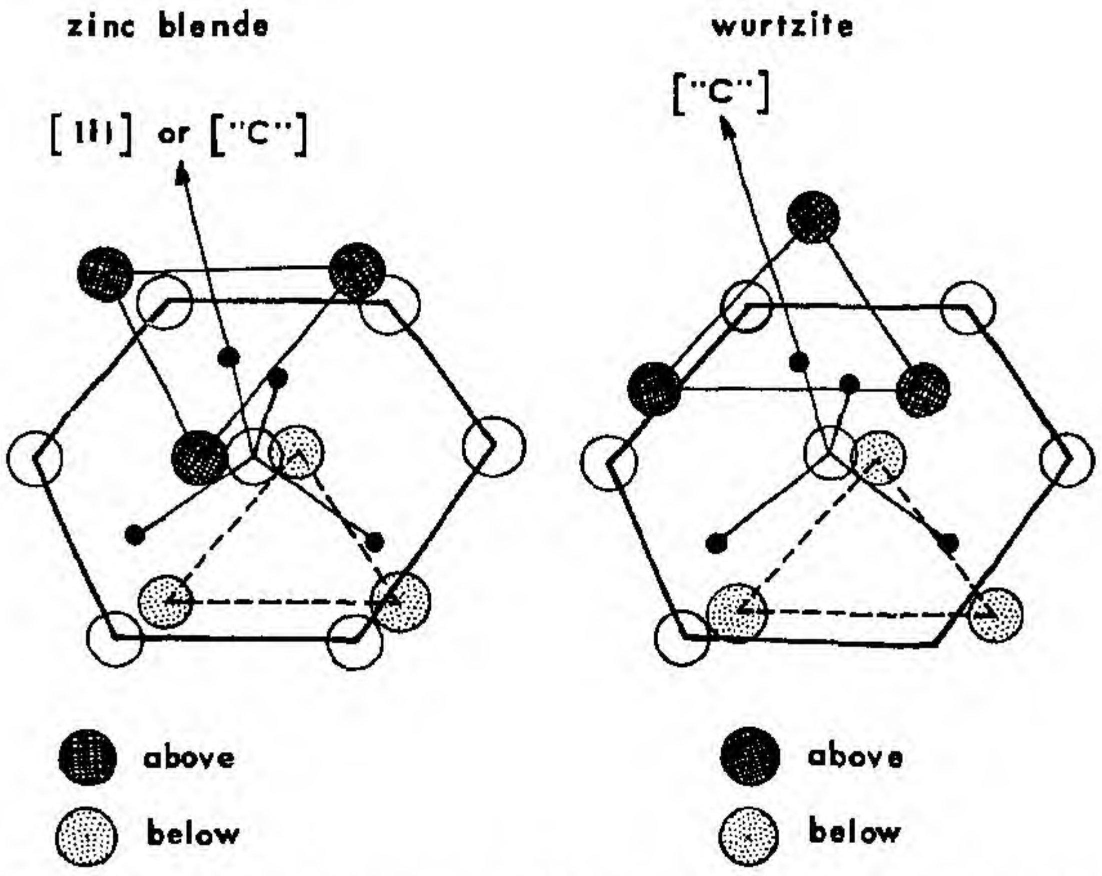


FIG. 1. First and second neighbors in zinc blende and wurtzite. Large circles are S atoms, small ones Zn. Open circles are in the same plane. Comparing the two structures, we note that only three of the twelve second neighbors differ, and even these are disposed symmetrically. These are shown as the 3 atoms "above" and are rotated by $\pi/3$ in zinc blende with respect to their positions in wurtzite.

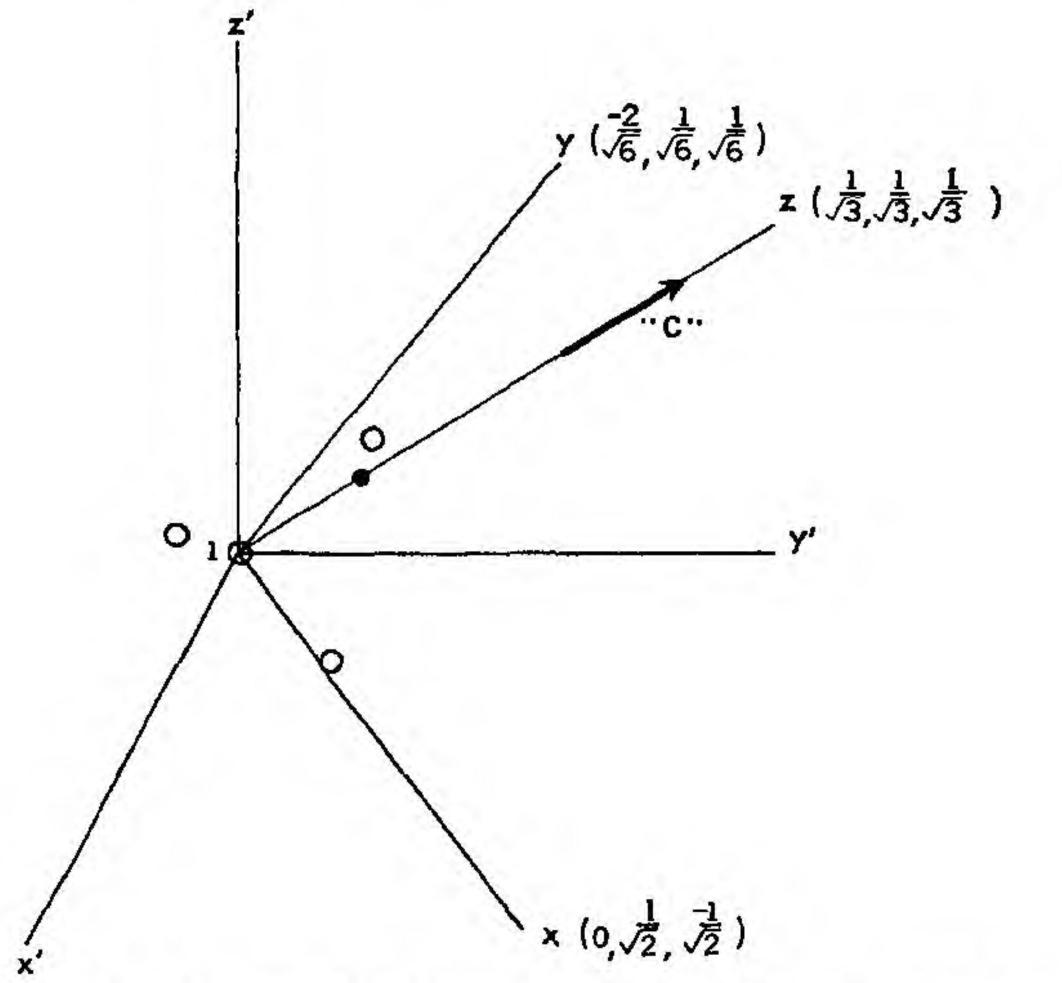


FIG. 2. Axes for zinc blende and wurtzite. The axes x, y, z are the ones used in this paper. These are the axes usually employed for the wurtzite structure. The axes x', y', z' are the conventional axes for the zinc-blende structure. The open and blackened circles represent sulfur and zinc sites, respectively.

jth cell in zinc blende. These functions form an orthonormal set ($\langle\Psi_n^{ZB}|\Psi_{n'}^{ZB}\rangle=\delta_{nn'}$) and are a complete cell-periodic set in zinc blende. In wurtzite (W), the Hamiltonian matrix at $\mathbf{k}=0$ will be computed using as basis functions the linear combinations of zinc-blende Wannier functions,

$$\Psi_n^{W\pm}=(2N)^{-\frac{1}{2}}\sum_j[\chi_n(\mathbf{r}-\mathbf{R}_j^1)\pm\chi_n(\mathbf{r}-\mathbf{R}_j^2)],$$

where $\mathbf{R}_j^1$ and $\mathbf{R}_j^2$ are the coordinates of the two sulfur atoms in the cell in wurtzite. We remark that $\langle\Psi_n^{W+}|\Psi_{n'}^{W-}\rangle=\langle\Psi_n^{W+}|H_W|\Psi_{n'}^{W-}\rangle=0$; in other words, the wurtzite Hamiltonian matrix breaks up into two submatrices spanned, respectively, by the functions Ψ_n^{W+} and Ψ_n^{W-}. This is proved by noting that the symmetry operation C_2, the twofold screw axis in wurtzite, which interchanges type-1 and type-2 sites, leaves H_W and the Ψ_n^{W+} functions invariant while changing the sign of the Ψ_n^{W-} functions. Furthermore, the functions Ψ_n^{W+} and Ψ_n^{W-} are similar in form, respectively, to the zinc-blende valence band wave functions $\Psi_{n\mathbf{k}}^{ZB}=N^{-\frac{1}{2}}\sum_j\exp(i\mathbf{k}\cdot\mathbf{R}_j)\chi_n(\mathbf{r}-\mathbf{R}_j)$ at the points $\mathbf{k}=(0,0,0)$ and $\mathbf{k}=(0,0,2\pi/c)$ in the zinc-blende Brillouin zone. Consequently, the wurtzite energy eigenvalues determined by the submatrix $\langle\Psi_n^{W+}|H_W|\Psi_{n'}^{W+}\rangle$ correspond to levels at $\mathbf{k}=(0,0,0)$ in zinc blende, while those determined by the submatrix $\langle\Psi_n^{W-}|H_W|\Psi_{n'}^{W-}\rangle$ correspond to levels at $\mathbf{k}=(0,0,2\pi/c)$, which is the point Λ at the Brillouin zone edge in zinc blende. Only the wurtzite levels corresponding to those at $\mathbf{k}=(0,0,0)$ in zinc blende will be considered.

To a good approximation, the functions Ψ_n^{W+} are an orthonormal set. The reason is that the nearest *like ion* (second-nearest neighbor) configurations in wurtzite and zinc blende are almost identical, and consequently nearest-like-ion overlaps can be expected to give roughly equal contributions to $\langle\Psi_n^{ZB}|\Psi_{n'}^{ZB}\rangle$ and $\langle\Psi_n^{W+}|\Psi_{n'}^{W+}\rangle$. Assuming higher overlaps can be neglected, we have $\langle\Psi_n^{W+}|\Psi_{n'}^{W+}\rangle\approx\langle\Psi_n^{ZB}|\Psi_{n'}^{ZB}\rangle=\delta_{nn'}$. Thus, by means of a linear combination of zinc-blende Wannier functions one can construct a nearly orthonormal set of basis functions for the wurtzite structure, suitable for a comparison of the energy levels in the two structures at $\mathbf{k}=(0,0,0)$.

Zinc Blende

Let us consider in more detail the zinc-blende Bloch functions at the top of the valence band. If we omit the spin-orbit interaction term in the zinc-blende Hamiltonian, the top of the valence band will be sixfold degenerate, with state vectors $(|X\rangle,|Y\rangle,|Z\rangle)\cdot(|+\rangle,|-\rangle)$ where $|X\rangle$, $|Y\rangle$, $|Z\rangle$ are Bloch states transforming like x, y, z under the operations of the zinc-blende symmetry group T_d. The zinc-blende Hamiltonian with the spin-orbit interaction term is diagonal in the manifold spanned by

$$\Gamma_8\begin{cases}|1\rangle=|\Pi\rangle|+\rangle,\\|2\rangle=(1/\sqrt{6})[\sqrt{2}|\Pi\rangle|-\rangle-2|Z\rangle|+\rangle],\\|3\rangle=(1/\sqrt{6})[\sqrt{2}|\bar{\Pi}\rangle|+\rangle+2|Z\rangle|-\rangle],\\|4\rangle=|\bar{\Pi}\rangle|-\rangle,\end{cases}\tag{1}$$

$$\Gamma_7\begin{cases}|5\rangle=(1/\sqrt{3})[\sqrt{2}|\Pi\rangle|-\rangle+|Z\rangle|+\rangle],\\|6\rangle=(1/\sqrt{3})[\sqrt{2}|\bar{\Pi}\rangle|+\rangle-|Z\rangle|-\rangle],\end{cases}$$

$$|\Pi\rangle=(1/\sqrt{2})(|X\rangle+i|Y\rangle),\quad|\bar{\Pi}\rangle=(1/\sqrt{2})(|X\rangle-i|Y\rangle).$$

The first group of states transforms as a basis for the irreducible representation Γ_8 of the double group of T_d; the second set is a basis for Γ_7. Use of the finite basis, Eq. (1), to diagonalize the Hamiltonian is equivalent to treating the spin-orbit term by first-order perturbation theory. Thus, the error in the energies made by the

neglect of admixtures of wave functions from other bands is of the order δ^2/E_g, where δ is the zinc-blende spin-orbit splitting and E_g is the band gap. Since $\delta^2/E_g \approx 0.067\delta/3.6$, the fractional error made in the valence band splitting is small. The prediction that the zinc-blende valence band is split into a Γ_8 level and a Γ_7 level agrees with experiment.[1]

Wurtzite

The basis functions in wurtzite are taken as sums of zinc-blende Wannier functions. Consider in particular those constructed from the zinc-blende valence band wave functions, using for these the approximate forms Eqs. (1). Since the threefold rotation operation and the reflection plane parallel to its axis are symmetry operations in *both* zinc blende and wurtzite, the behavior of the sets of states $|1\rangle, \cdots, |6\rangle$; $|1^+\rangle, \cdots, |6^+\rangle$ under these operations will be identical. It is thus possible to show, by using characters of the double group of C_{3v}, that the states $|1^+\rangle, \cdots, |6^+\rangle$ transform according to representations of the double group of C_{3v} as follows[9]

$$\begin{array}{lll} \begin{matrix}\Lambda_4 & (1/\sqrt{2})[|1^+\rangle+|4^+\rangle] \\ \Lambda_5 & (1/\sqrt{2})[|1^+\rangle-|4^+\rangle] \\ (\Gamma_9) & \end{matrix} & \begin{matrix}\Lambda_6\begin{cases}|2^+\rangle \\ |3^+\rangle\end{cases} \\ (\Gamma_7)\end{matrix} & \begin{matrix}\Lambda_6\begin{cases}|5^+\rangle \\ |6^+\rangle\end{cases} \\ (\Gamma_7)\end{matrix} \end{array}.$$

This permits a simplification of the submatrix of the wurtzite Hamiltonian spanned by $|1^+\rangle, \cdots, |6^+\rangle$, to

	1^+	4^+	2^+	5^+	3^+	6^+
1^+	a	0	0	0	0	0
4^+	0	a	0	0	0	0
2^+	0	0	b	c	0	0
5^+	0	0	c^*	d	0	0
3^+	0	0	0	0	b	c
6^+	0	0	0	0	c^*	d

The prediction that the wurtzite valence band consists of a Γ_9 level and two Γ_7 levels agrees with experiment.

Just as in zinc blende, the admixture of other wave functions forming the basis will be neglected. Let us write

$$\langle n|H_0|n'\rangle = \langle \Psi_n^{\mathrm{ZB}}|H_{\mathrm{ZB}}|\Psi_{n'}^{\mathrm{ZB}}\rangle - \delta\delta_{nn'}\Big(\sum_{k=1}^{4}\delta_{nk}\Big),$$

$$\langle n|H_1|n'\rangle = \langle \Psi_n^{\mathrm{W}+}|H_{\mathrm{W}}|\Psi_{n'}^{\mathrm{W}+}\rangle - \langle \Psi_n^{\mathrm{ZB}}|H_{\mathrm{ZB}}|\Psi_{n'}^{\mathrm{ZB}}\rangle + \delta\delta_{nn'}\Big(\sum_{k=1}^{4}\delta_{nk}\Big),$$

[9] The designation in parentheses refers to the representation of the double group of C_{6v} according to which the functions approximately transform. The basis functions for wurtzite would transform exactly according to irreducible representations of C_{6v} if we took the zinc-blende valence band Bloch functions to be linear combinations of ionic $3p$ orbitals. We have instead used the exact zinc-blende Bloch functions. Because the wurtzite Hamiltonian is diagonalized with respect to only a finite basis, and because the functions in the basis transform only approximately according to irreducible representations of C_{6v}, the wurtzite valence band wave functions obtained have only approximately the correct symmetry properties. Nevertheless, in the remainder of this paper the wurtzite functions will be labeled by the indicated representations of C_{6v}.

where δ, as before, is the zinc-blende spin-orbit splitting, and $\delta_{nn'}$ is the Kronecker delta. The near degeneracy of the valence band in zinc blende has been made a complete degeneracy in H_0 by subtraction of $\delta\delta_{nn'}(\sum_{k=1}^{4}\delta_{nk})$; this term has been included in the perturbation H_1. In this way we have defined a problem in degenerate perturbation theory. If the degenerate manifold is treated exactly, neglect of the admixture of other wave functions of the basis is equivalent to neglecting second- and higher order terms in a perturbation expansion, and results in an error $\approx|\langle|H_1|\rangle|^2/E_g$. Here $|\langle|H_1|\rangle|$ is the magnitude characteristic of matrix elements of the perturbation, which can be expected to be of the order of the valence band splittings in wurtzite ≈ 0.05 ev. Thus, the fractional error made in the valence band splitting $\approx 0.05/3.6$, which is small.

Two-Parameter Formulas

Because H_1 factors completely into 2×2 submatrices, exact solution of the eigenvalue problem within the degenerate manifold is easy. The results for the wurtzite and zinc-blende valence band energy splittings and the wurtzite valence band eigenstates are:

$$\Delta E_{\mathrm{W}}:\quad \begin{cases} E_a - E_b = \dfrac{\alpha+\delta}{2} - \left[\left(\dfrac{\alpha+\delta}{2}\right)^2 - \dfrac{2\alpha\delta}{3}\right]^{\frac{1}{2}}, \\[2ex] E_b - E_c = 2\left[\left(\dfrac{\alpha+\delta}{2}\right)^2 - \dfrac{2\alpha\delta}{3}\right]^{\frac{1}{2}}, \end{cases} \tag{2}$$

$$\Delta E_{\mathrm{ZB}}:\qquad \delta = \langle 1|H_{\mathrm{ZB}}|1\rangle - \langle 5|H_{\mathrm{ZB}}|5\rangle;$$

$$|a\rangle = |1^+\rangle,$$

$$\begin{aligned} N_b|b\rangle &= \frac{\sqrt{2}\alpha}{3}|5^+\rangle + \left\{\frac{\delta}{2} - \frac{\alpha}{6} + \left[\left(\frac{\alpha+\delta}{2}\right)^2 - \frac{2\alpha\delta}{3}\right]^{\frac{1}{2}}\right\}|2^+\rangle, \\ N_c|c\rangle &= \frac{\sqrt{2}\alpha}{3}|5^+\rangle + \left\{\frac{\delta}{2} - \frac{\alpha}{6} - \left[\left(\frac{\alpha+\delta}{2}\right)^2 - \frac{2\alpha\delta}{3}\right]^{\frac{1}{2}}\right\}|2^+\rangle. \end{aligned} \tag{3}$$

Here N_b and N_c are normalization constants and α is a crystal field parameter defined by

$$\alpha = [\langle \Pi^+|H_{\mathrm{W}}|\Pi^+\rangle - \langle \Pi|H_{\mathrm{ZB}}|\Pi\rangle] - [\langle Z^+|H_{\mathrm{W}}|Z^+\rangle - \langle Z|H_{\mathrm{ZB}}|Z\rangle].$$

In this expression H_{W} and H_{ZB} signify the Hamiltonians without the spin-orbit interaction terms.[10]

Equations (2) are the two-parameter formulas originally derived by Hopfield.[5] They have been obtained by making only three approximations:

[10] It is easy to show that the fractional error resulting from neglect of these terms is of the order

$$\frac{|\langle 1^+|(\hbar/4m^2c^2)\nabla V_{\mathrm{W}}\times\mathbf{p}|1^+\rangle|}{|\langle 1^+|V_{\mathrm{W}}|1^+\rangle|} \approx \frac{\hbar^2}{m^2c^2\langle r^2\rangle} \approx 10^{-4},$$

where $\langle r^2\rangle$ is the mean square sulfur ion radius.

(1) Assumption of approximate orthonormality of the wurtzite basis;
(2) Neglect of energy terms $\approx\delta^2/E_g$;
(3) Neglect of energy terms $\approx|\langle|H_1|\rangle|^2/E_g$.

Note that the equations for the splittings in wurtzite are completely symmetric in α and δ. This means that α and δ cannot be determined uniquely from the wurtzite splittings: if $(\alpha,\delta)=(a,b)$ is one solution consistent with the data, then $(\alpha,\delta)=(b,a)$ is another. The ambiguity just corresponds to the fact that solving (2) for α and δ in terms of E_a-E_b, E_b-E_c leads to a quadratic equation, both roots of which are allowable solutions. This symmetry of the two-parameter formulas is not explicitly evident in the version of them given by Balkanski and Cloizeux.[11]

Finally, it is to be emphasized that the anion and cation have been assumed to be S and Zn in the above derivation purely for the sake of convenience in referring to them. The theory should be valid in other semiconductors with p-like valence bands and with spin-orbit and crystal field splittings which are small relative to the band gap.

3. APPLICATION TO ZnS, CdS, AND OTHER II-VI COMPOUNDS

The two-parameter formulas describe three splittings in terms of two parameters. They fit well the values given by Birman *et al.*[1] for the spin-orbit splitting of the zinc-blende form and the two valence band splittings of the wurtzite form of ZnS. From the data (Table I) at 77°K, $\delta=0.068$ ev and $E_a-E_b+\frac{1}{2}(E_b-E_c)=\frac{1}{2}(\alpha+\delta)=0.069$ ev, giving $\alpha=0.070$ ev. The theory then gives 0.080 ev and 0.029 ev for the wurtzite splittings, within 10% of the experimental values of 0.084 ev and 0.027 ev. The order of levels predicted is also correct.

A reasonable result is also obtained when Eqs. (2) are applied to data for CdS. Crystals of cubic CdS have not yet been grown. However, with data for hexagonal CdS, the formulas can be used to predict the value of the spin-orbit splitting which would be observed in cubic CdS. Two values are obtained as a result of the ambiguity discussed above: $\delta=0.065$ ev or $\delta=0.029$ ev (at 77°K). The first of these is close to the ZnS splitting and is probably the correct solution, since the valence band spin-orbit splitting should be determined primarily by the wave function and potential near the sulfur ion and should depend only weakly on the nature of the cation. Measurements on CdSe[2] and on ZnSe,[4] which show nearly the same spin-orbit splitting for both substances, are evidence for the validity of this type of argument.

TABLE I. Valence band splittings in ZnS and CdS.

Temperature (°K)	δ (ev)	E_a-E_b (ev)	E_b-E_c (ev)
	Cubic ZnS		
77	0.068		
14	0.065		
	Hexagonal ZnS		
77		0.027	0.084
14		0.026	0.082
	Hexagonal CdS		
77		0.016	0.062

[11] M. Balkanski and J. Cloizeux (to be published).

From the formulas, Eqs. (3), the parentage of the lines in wurtzite ZnS can be determined. Taking $\alpha=0.070$ ev, one obtains

$$|b\rangle=0.48|5^+\rangle+0.88|2^+\rangle,$$
$$|c\rangle=0.88|5^+\rangle-0.48|2^+\rangle.$$

These can also be written

$$|b\rangle=0.90|\Pi^+\rangle|-\rangle-0.44|Z^+\rangle|+\rangle,$$
$$|c\rangle=0.44|\Pi^+\rangle|-\rangle+0.90|Z^+\rangle|+\rangle.$$

These expressions have a simple interpretation (see the splitting diagram, Fig. 3). The states $|5\rangle$ and $|2\rangle$ transform according to the irreducible representations Γ_5 and Γ_1, respectively, of the double group of the wave vector at $\mathbf{k}=(0,0,0)$ in zinc blende. These states are zinc-blende valence band eigenstates which differ in energy by δ, the spin-orbit energy. Equations (3) indicate how these zinc-blende levels mix when the crystal field perturbation is "turned on." The states $|\Pi^+\rangle|-\rangle$ and $|Z^+\rangle|+\rangle$ transform, respectively, according to the irreducible representations Γ_5 and Γ_1 of the single group C_{6v}, the

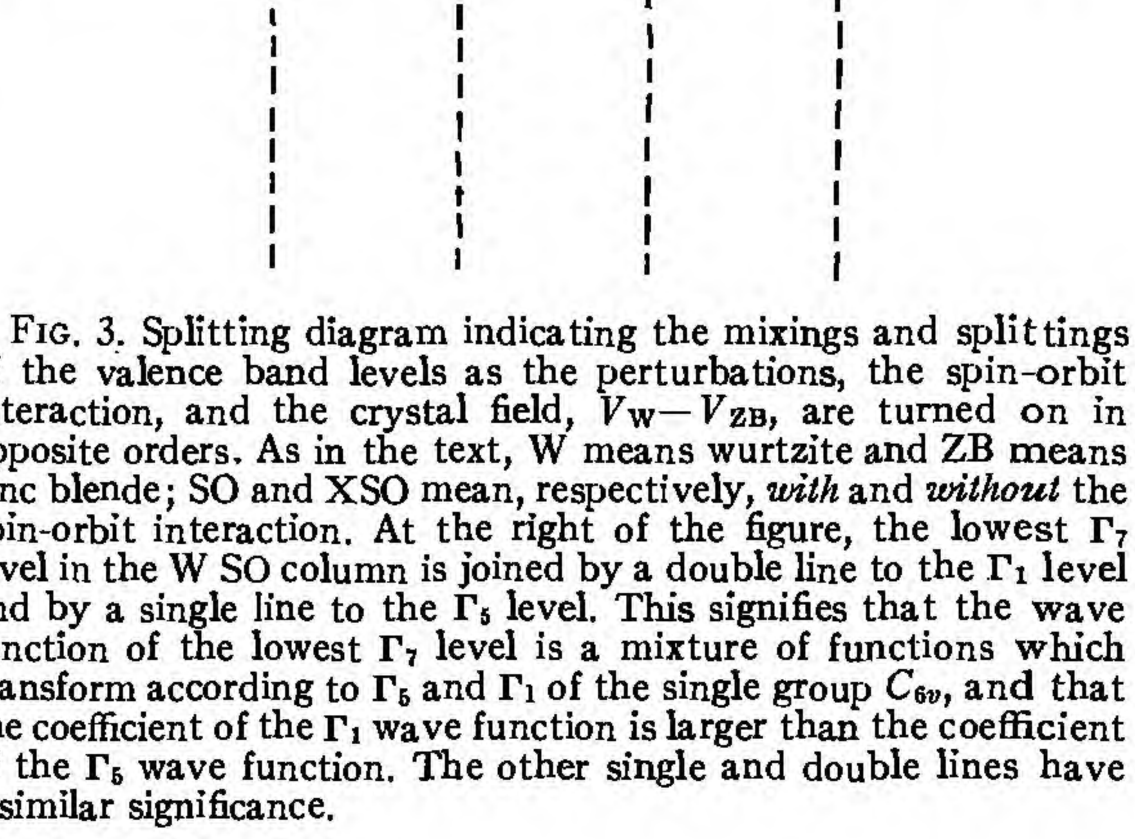


FIG. 3. Splitting diagram indicating the mixings and splittings of the valence band levels as the perturbations, the spin-orbit interaction, and the crystal field, V_W-V_{ZB}, are turned on in opposite orders. As in the text, W means wurtzite and ZB means zinc blende; SO and XSO mean, respectively, *with* and *without* the spin-orbit interaction. At the right of the figure, the lowest Γ_7 level in the W SO column is joined by a double line to the Γ_1 level and by a single line to the Γ_5 level. This signifies that the wave function of the lowest Γ_7 level is a mixture of functions which transform according to Γ_5 and Γ_1 of the single group C_{6v}, and that the coefficient of the Γ_1 wave function is larger than the coefficient of the Γ_5 wave function. The other single and double lines have a similar significance.

group of the wave vector at $\mathbf{k}=(0,0,0)$ in a wurtzite structure in which there is no spin-orbit interaction. According to Eqs. (3), when the spin-orbit interaction in wurtzite is "turned off" by setting $\delta=0$, the state $|b\rangle$ becomes $|\Pi^+\rangle|-\rangle$ (transforming according to Γ_5), the state $|c\rangle$ becomes $|Z^+\rangle|+\rangle$ (transforming according to Γ_1), and the Γ_5 state has the higher energy. The prediction that the Γ_5 level would exceed the Γ_1 level in energy in wurtzite ZnS if the spin-orbit interaction could be eliminated agrees with the observations showing that in hexagonal ZnO, in which the spin-orbit interaction is extremely weak due to the low anion atomic number, the Γ_5-like levels lie above a Γ_1-like level.[12]

4. ESTIMATION OF THE EFFECTIVE CHARGE

An attempt was made to make an *a priori* calculation of α, and by this means to estimate an effective charge for ZnS. The zinc-blende Bloch states $|X\rangle$, $|Y\rangle$, $|Z\rangle$ were assumed to be LCAO states constructed from sulfur $3p$ ionic orbitals. Only zeroth and first-neighbor interaction integrals were retained, and a point-ion model of the zinc-blende and wurtzite lattices, with effective ionic charges $+\lambda e$ and $-\lambda e$ for zinc and sulfur ions, respectively, was used. This model gives $\alpha=-(3E/5)\times\int_0^\infty dr\, r^4(R_{3p})^2$, where E is the coefficient of $r^2P_2(\cos\theta)$ in the expansion of V_W about a sulfur ion site and R_{3p} is an ionic radial orbital. Evaluation of E by an Ewald summation method[13] gives $\lambda=2.3$, clearly too large. Taking into account the effect of mixings of sulfur $3d$ states into the zinc-blende valence band wave function and the deviation of wurtzite ZnS from ideality was found to produce little change in the value of λ obtained. (The expansions of V_W and V_{ZB} used are given in the Appendix.) Mixings of zinc $4p$ states into the zinc-blende wave function may have an important effect on λ, but the overlap integrals necessary to estimate this were not calculated.

[12] However, in ZnO, unlike ZnS, the order of levels is Γ_7, Γ_9, Γ_7; this is discussed by Hopfield[6] and by Thomas.[3]

[13] B. R. A. Nijboer and F. W. De Wette, Physica 23, 309 (1957).

ACKNOWLEDGMENT

The author wishes to express his gratitude to Dr. J. L. Birman, who suggested the problem discussed in this paper and with whom the writer has had many valuable discussions.

APPENDIX

We give here the first few terms of the expansion in solid harmonics of the potentials in ideal wurtzite and zinc-blende point-ion lattices. The nearest like-ion distance is d_{SS} and the effective charge parameter is λ. The origin for the expansion is a sulfur site such as is at the orgin in Fig. 2; in other words, a sulfur site with nearest neighbors with direction cosines $(0,0,1)$; $(0, -2\sqrt{2}/3, -1/3)$; $(\sqrt{2}/\sqrt{3}, \sqrt{2}/3, -1/3)$; $(-\sqrt{2}/\sqrt{3}, \sqrt{2}/3, -1/3)$ on the unprimed axes. All sums were evaluated by an Ewald method and were checked to the number of places indicated by summing with two different values of the convergence parameter.

Zinc blende (conventional axes, i.e., primed axes in Fig. 2):

$$\frac{V_{ZB}}{e^2\lambda/d_{SS}}=A_0+B_0\frac{x'y'z'}{d_{SS}^3}+\cdots,$$

$$B_0=76.8.$$

Wurtzite (unprimed axes in Fig. 2):

$$\frac{V_W}{e^2\lambda/d_{SS}}=C_0+D_0\frac{z}{d_{SS}}+E_0\frac{\frac{1}{2}(2z^2-x^2-y^2)}{d_{SS}^2}+F_0\frac{\frac{1}{2}[2z^3-3(x^2+y^2)z]}{d_{SS}^3}+G_0\frac{(3x^2y-y^3)}{d_{SS}^3}+\cdots,$$

$$D_0=-0.0397,\quad F_0=14.43,$$

$$E_0=\ 0.142,\quad G_0=10.1.$$

Note that $E=(-e^2\lambda/d_{SS}^3)E_0$.

PHYSICAL REVIEW VOLUME 126, NUMBER 2 APRIL 15, 1962

Quantum Theory of the Dielectric Constant in Real Solids

STEPHEN L. ADLER*
General Electric Research Laboratory, Schenectady, New York

(Received October 12, 1961)

The quantum theory of the frequency- and wave-number-dependent dielectric constant in solids is extended in order to study the full dielectric constant tensor and to include local-field effects. Within the framework of the band theory, an explicit expression for the dielectric constant tensor, neglecting local-field effects, is derived. In addition to components which are the ordinary longitudinal and transverse dielectric constants, there are components which couple transverse and longitudinal electromagnetic disturbances. A formalism for calculating the local-field corrections to the dielectric constant is developed in detail for the case of the longitudinal dielectric constant of a cubic insulating solid. In the coarsest (dipole) approximation, the theory gives a Lorenz-Lorentz formula modified by self-polarization corrections arising from the polarization of the charge in a unit cell by its own field.

QUANTUM mechanical treatments of the frequency-and wave-number-dependent dielectric constant in solids have been given by Nozières and Pines[1] and by Ehrenreich and Cohen.[2] These authors give explicit expressions for certain components of the dielectric constant tensor, valid within the framework of the random phase approximation (RPA). Expressions are not given for the remaining components of the dielectric constant tensor and local field effects are neglected. This paper will generalize the treatment of Ehrenreich and Cohen[2] so as to include additional effects of interest in real solids. In Sec. I an expression for the full frequency- and wave-number-dependent dielectric constant tensor in a solid of arbitrary symmetry will be derived, still neglecting local field effects. The additional components obtained correspond to a coupling between longitudinal and transverse electromagnetic disturbances. This coupling, which does not appear in an isotropic free electron gas, is present in solids of even cubic symmetry and vanishes only for propagation along special directions of high symmetry. In Sec. II, local field effects in insulators of cubic symmetry will be discussed. An integral equation will be set up which determines the longitudinal dielectric constant with local field corrections in the case of wavelengths large compared to the lattice constant but small compared to the over-all crystal dimensions. The integral equation will be rewritten by making a multipole expansion of the potential in a given cell arising from the charge density in all other cells. The solution, when only dipole terms are retained, is the modified Lorenz-Lorentz formula

$$\epsilon-1=\frac{4\pi(\alpha-C_1)}{1-(4\pi/3)(\alpha-C_1)},$$

where α is the polarizability of the solid calculated without making local field corrections, and C_1 is a reduction in α due to polarization of the charge in a given cell by its own field.

The calculations of Secs. I and II are performed within the framework of the one-electron (energy band) approximation and use a linearized Liouville equation to determine the single-particle density matrix. Since in this context linearization is equivalent to the RPA,[2] the results obtained in Secs. I and II are still valid only within the framework of the RPA.

I. DIELECTRIC CONSTANT TENSOR

We will first introduce a phenomenological dielectric constant tensor. Let $\mathbf{A}(\mathbf{r},t)$ and $\phi(\mathbf{r},t)$ be the potentials describing fields acting on a system of charged particles. In response to the fields, charge and current densities $\rho^{\text{ind}}(\mathbf{r},t)$ and $\mathbf{j}^{\text{ind}}(\mathbf{r},t)$, which satisfy the equation of continuity $\nabla\cdot\mathbf{j}^{\text{ind}}+\partial\rho^{\text{ind}}/\partial t=0$, will be induced in the system. Let us immediately introduce Fourier transforms $\mathbf{A}(\mathbf{q},\omega)$, $\phi(\mathbf{q},\omega)$, $\mathbf{j}^{\text{ind}}(\mathbf{q},\omega)$, etc., by

$$\mathbf{A}(\mathbf{r},t)=\int d\mathbf{q}d\omega\,\mathbf{A}(\mathbf{q},\omega)\exp i(\mathbf{q}\cdot\mathbf{r}-\omega t), \tag{1.1}$$

and similar equations.

In their treatments of the dielectric constant, Nozières and Pines[1] and Ehrenreich and Cohen,[2] following a practice originated by Lindhard,[3] define longitudinal and transverse dielectric constants $\boldsymbol{\varepsilon}^L$ and $\boldsymbol{\varepsilon}^T$ by the equations

$$-i\omega[\boldsymbol{\varepsilon}^{(L,T)}(\mathbf{q},\omega)-1]\cdot\mathbf{E}^{(L,T)}(\mathbf{q},\omega)=4\pi\mathbf{j}^{\text{ind}(L,T)}(\mathbf{q},\omega). \tag{1.2}$$

The two constants describe respectively the longitudinal current induced by a purely-longitudinal electric field and the transverse current induced by a purely-transverse electric field. In the case of a free-electron gas a longitudinal (transverse) current cannot be induced by a transverse (longitudinal) electric field. Consequently $\boldsymbol{\varepsilon}^L$ and $\boldsymbol{\varepsilon}^T$ give a complete description of the linear dielectric properties. In solids, in general a

* Present address: Department of Physics, Princeton University, Princeton, New Jersey.
[1] P. Nozières and D. Pines, Phys. Rev. **109**, 741, 762, 1062 (1958); *ibid.* **111**, 442 (1958); *ibid.* **113**, 1254 (1959).
[2] H. Ehrenreich and M. H. Cohen, Phys. Rev. **115**, 786 (1959).
[3] J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **28**, 8 (1954).

purely-transverse or a purely-longitudinal electric field induces both transverse and longitudinal currents. In this case the linear dielectric properties are fully described by a dielectric-constant tensor defined by

$$-i\omega[\boldsymbol{\varepsilon}(\mathbf{q},\omega)-\mathbf{1}]\cdot\mathbf{E}(\mathbf{q},\omega)=4\pi\mathbf{j}^{\text{ind}}(\mathbf{q},\omega). \tag{1.3}$$

The longitudinal and transverse constants $\varepsilon^L(\mathbf{q},\omega)$ and $\varepsilon^T(\mathbf{q},\omega)$ can be simply related to $\boldsymbol{\varepsilon}(\mathbf{q},\omega)$. Let $\hat{q}$ be a unit vector parallel to the direction of propagation $\mathbf{q}$, and define

$$\begin{aligned}\mathbf{1}_L&=\hat{q}\hat{q}\\ \mathbf{1}_T&=\mathbf{1}-\hat{q}\hat{q}\end{aligned} \tag{1.4}$$

where $\mathbf{1}$ is the unit dyadic. Then letting $\mathbf{E}(\mathbf{q},\omega)$ be purely longitudinal or purely transverse gives

$$\begin{aligned}\varepsilon^L(\mathbf{q},\omega)&=\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L,\\ \varepsilon^T(\mathbf{q},\omega)&=\mathbf{1}_T\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_T.\end{aligned} \tag{1.5}$$

The remaining components of the dielectric-constant tensor are $\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_T$ and $\mathbf{1}_T\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L$, which vanish for a free-electron gas but do not in general vanish for a solid. They describe, respectively, the longitudinal ttransverse) current induced by a transverse (longi-(udinal) electric field.

An explicit expression for $\boldsymbol{\varepsilon}(\mathbf{q},\omega)$ will be calculated in the energy-band approximation. Consider the single-particle Liouville equation

$$i\hbar\partial\rho/\partial t=[H,\rho], \tag{1.6}$$

where ρ is the single-particle density matrix and

$$H=(1/2m)[\mathbf{p}-(e/c)\mathbf{A}(\mathbf{r},t)]^2+e\phi(\mathbf{r},t)+U(\mathbf{r}). \tag{1.7}$$

Here $U(\mathbf{r})$ is the periodic lattice potential. Let the state functions for the unperturbed lattice be $|\mathbf{k}l\rangle=V^{-\frac{1}{2}}u_{\mathbf{k}l}\times\exp(i\mathbf{k}\cdot\mathbf{r})$ with $u_{\mathbf{k}l}$ cell-periodic and V the volume of the crystal. They satisfy the Schrödinger equation $[\mathbf{p}^2/2m+U(\mathbf{r})]|\mathbf{k}l\rangle=E_{\mathbf{k}l}|\mathbf{k}l\rangle$, in which $\mathbf{k}$ is the wave vector and l the band index. Linearize the Liouville equation with respect to $\mathbf{A}$ and ϕ by setting $\rho=\rho^{(0)}+\rho^{(1)}$ etc. The unperturbed density matrix, $\rho^{(0)}$, satisfies $\rho^{(0)}|\mathbf{k}l\rangle=f_0(E_{\mathbf{k}l})|\mathbf{k}l\rangle$ with $f_0(E_{\mathbf{k}l})$ the Fermi-Dirac distribution function. The perturbation $\rho^{(1)}$ is linear in $\mathbf{A}$ and ϕ. Dropping quadratic terms gives

$$\begin{aligned}i\hbar\partial\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle/\partial t&\\ =(E_{l'\mathbf{k}+\mathbf{q}}-E_{l\mathbf{k}})&\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle\\ +[f_0(E_{l\mathbf{k}})&-f_0(E_{l'\mathbf{k}+\mathbf{q}})]\\ &\times\langle l'\mathbf{k}+\mathbf{q}|-(e/2mc)(\mathbf{A}\cdot\mathbf{p}+\mathbf{p}\cdot\mathbf{A})+e\phi|l\mathbf{k}\rangle.\end{aligned} \tag{1.8}$$

Let us assume

$$\begin{aligned}\mathbf{A}(\mathbf{r},t)&=\mathbf{A}(\mathbf{q},\omega)\exp i(\mathbf{q}\cdot\mathbf{r}-\omega t),\\ \phi(\mathbf{r},t)&=\phi(\mathbf{q},\omega)\exp i(\mathbf{q}\cdot\mathbf{r}-\omega t),\end{aligned} \tag{1.9}$$

and make the Ansatz that the time dependence of $\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle$ is $\exp(-i\omega t)$. The frequency ω is taken to have a small positive imaginary part, corresponding to an adiabatic turning on of the perturbing potentials. It is an easy calculation to show that

$$\langle l'\mathbf{k}+\mathbf{q}'|e\phi|l\mathbf{k}\rangle=\delta_{\mathbf{q}',\mathbf{q}}e\phi(\mathbf{q},\omega)(l'\mathbf{k}+\mathbf{q}|l\mathbf{k}) \tag{1.10}$$

and

$$\begin{aligned}\tfrac{1}{2}\langle l'\mathbf{k}+\mathbf{q}'|\mathbf{A}\cdot\mathbf{p}+\mathbf{p}\cdot\mathbf{A}|l\mathbf{k}\rangle&\\ =\delta_{\mathbf{q}',\mathbf{q}}(l'\mathbf{k}+\mathbf{q}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l\mathbf{k})&\cdot\mathbf{A}(\mathbf{q},\omega).\end{aligned} \tag{1.11}$$

The abbreviation

$$\begin{aligned}(l'\mathbf{k}+\mathbf{q}|f(\mathbf{r}_e,\mathbf{p}_e)|l\mathbf{k})&\\ \equiv(1/v_a)\int_0 u_{l'\mathbf{k}+\mathbf{q}}{}^*(\mathbf{r})f(\mathbf{r},\,-i\hbar\nabla_r)u_{l\mathbf{k}}(\mathbf{r})d\mathbf{r},&\end{aligned} \tag{1.12}$$

has been introduced, in which the integration extends over a unit cell. Couplings of the wave vector $\mathbf{q}$ to wave vectors $\mathbf{q}+\mathbf{K}$, where $\mathbf{K}$ is a reciprocal lattice vector, have been neglected. These so-called Umklapp processes give rise to the local field corrections and will be discussed in Sec. II. The solution of Eq. (1.8) is immediately obtained in the form

$$\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle=\frac{[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})][(l'\mathbf{k}+\mathbf{q}|l\mathbf{k})e\phi(\mathbf{q},\omega)-(l'\mathbf{k}+\mathbf{q}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l\mathbf{k})\cdot(e/mc)\mathbf{A}(\mathbf{q},\omega)]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}. \tag{1.13}$$

The induced current and charge density may be calculated from

$$\begin{aligned}\mathbf{j}^{\text{ind}}(\mathbf{r},t)&=\mathrm{Tr}\rho^{(1)}\mathbf{j}_{\text{op}}{}^{(0)}(\mathbf{r})+\mathrm{Tr}\rho^{(0)}\mathbf{j}_{\text{op}}{}^{(1)}(\mathbf{r},t),\\ \rho^{\text{ind}}(\mathbf{r},t)&=\mathrm{Tr}\rho^{(1)}\rho_{\text{op}}{}^{(0)}(\mathbf{r}),\end{aligned} \tag{1.14}$$

where

$$\begin{aligned}\mathbf{j}_{\text{op}}{}^{(0)}(\mathbf{r})&=(\hbar/2m)[(\mathbf{p}_e/m)\delta(\mathbf{r}-\mathbf{r}_e)+\delta(\mathbf{r}-\mathbf{r}_e)(\mathbf{p}_e/m)],\\ \mathbf{j}_{\text{op}}{}^{(1)}(\mathbf{r})&=-(e/mc)\mathbf{A}(\mathbf{r},t)\delta(\mathbf{r}-\mathbf{r}_e),\\ \rho_{\text{op}}{}^{(0)}(\mathbf{r})&=e\delta(\mathbf{r}-\mathbf{r}_e),\end{aligned} \tag{1.15}$$

and $\mathbf{r}_e$ and $\mathbf{p}_e$ are, respectively, the position and momentum operators. This gives

$$\begin{aligned}\mathbf{j}^{\text{ind}}(\mathbf{q},\omega)=&-e^2\mathbf{A}(\mathbf{q},\omega)N/mcV\\ &+\textstyle\sum_{ll'\mathbf{k}}(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})\\ &\times\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle,\end{aligned} \tag{1.16}$$

$$\begin{aligned}\rho^{\text{ind}}(\mathbf{q},\omega)=(e/V)&\textstyle\sum_{ll'\mathbf{k}}(l\mathbf{k}|l'\mathbf{k}+\mathbf{q})\\ &\times\langle l'\mathbf{k}+\mathbf{q}|\rho^{(1)}|l\mathbf{k}\rangle,\end{aligned} \tag{1.17}$$

with N the number of cells in the crystal.

Since $\mathbf{j}^{\text{ind}}(\mathbf{q},\omega)$ and $\rho^{\text{ind}}(\mathbf{q},\omega)$ are obtained by linearization of a gauge-invariant theory with respect to the potentials, they are invariant under infinitesimal gauge transformations, and thus also under arbitrary gauge

transformations. This is verified explicitly, in the case of the expression for $\mathbf{j}^{\text{ind}}(\mathbf{q},\omega)$, in the Appendix. Since the theory is gauge invariant, we may transform to a gauge in which $\phi=0$ without loss of generality. Note that $\mathbf{A}$ will not in general be transverse in this gauge. From Eqs. (1.13) and (1.16) we get

$$\mathbf{j}^{\text{ind}}(\mathbf{q},\omega)=-e^2\mathbf{A}(\mathbf{q},\omega)N/mcV$$
$$-\frac{e^2}{m^2cV}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})](l'\mathbf{k}+\mathbf{q}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l\mathbf{k})\cdot\mathbf{A}(\mathbf{q},\omega)}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}. \quad (1.18)$$

Comparing with Eq. (1.3) and noting that $\mathbf{E}(\mathbf{q},\omega)=i\omega\mathbf{A}(\mathbf{q},\omega)/c$ gives

$$\boldsymbol{\varepsilon}(\mathbf{q},\omega)=(1-4\pi e^2N/mV\omega^2)\mathbf{1}$$
$$+\frac{4\pi e^2}{m^2V\omega^2}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})[f_0(E_{l'\mathbf{k}+\mathbf{q}})-f_0(E_{l\mathbf{k}})](l'\mathbf{k}+\mathbf{q}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l\mathbf{k})}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}, \quad (1.19)$$

an explicit expression for the frequency- and wave-number-dependent dielectric constant tensor.

It is interesting to compute $\varepsilon^L(\mathbf{q},\omega)=\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L$ directly from (1.19). This can be done by using the three identities

$$\mathbf{q}\cdot(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})$$
$$=(m/\hbar)(E_{l'\mathbf{k}+\mathbf{q}}-E_{l\mathbf{k}})(l\mathbf{k}|l'\mathbf{k}+\mathbf{q}), \quad (1.20)$$

$$0=-N+(1/\hbar q^2)\sum_{ll'\mathbf{k}}\mathbf{q}\cdot(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})$$
$$\times[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})](l'\mathbf{k}+\mathbf{q}|l\mathbf{k}), \quad (1.21)$$

and

$$0=\sum_{ll'\mathbf{k}}|(l\mathbf{k}|l'\mathbf{k}+\mathbf{q})|^2[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})]. \quad (1.22)$$

Equations (1.20) and (1.21) are proved in the Appendix, while Eq. (1.22) is obtained by making the change of variable $\mathbf{k}\to-\mathbf{k}-\mathbf{q}$ and noting that $E_{\mathbf{k}}=E_{-\mathbf{k}}$, $u_{l\mathbf{k}}=u_{l-\mathbf{k}}^*$, $E_{\mathbf{k}+\mathbf{K}}=E_{\mathbf{k}}$ and $u_{l\mathbf{k}+\mathbf{K}}=u_{l\mathbf{k}}$, where $\mathbf{K}$ is a vector of the reciprocal lattice. Applying the identities to (1.19) gives

$$\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L=\mathbf{1}_L\Bigg\{1-4\pi e^2N/mV\omega^2+\frac{4\pi e^2}{\hbar q^2mV\omega^2}\sum_{ll'\mathbf{k}}\mathbf{q}\cdot(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q})$$
$$\times[f_0(E_{l'\mathbf{k}+\mathbf{q}})-f_0(E_{l\mathbf{k}})](l'\mathbf{k}+\mathbf{q}|l\mathbf{k})\left(\frac{\hbar\omega}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}-1\right)\Bigg\}$$
$$=\mathbf{1}_L\left\{1+\frac{4\pi e^2}{q^2V}\sum_{ll'\mathbf{k}}\frac{|(l\mathbf{k}|l'\mathbf{k}+\mathbf{q})|^2[f_0(E_{l'\mathbf{k}+\mathbf{q}})-f_0(E_{l\mathbf{k}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}\right\}. \quad (1.23)$$

Equation (1.23) is just the longitudinal dielectric constant derived by Ehrenreich and Cohen.[2] Using the same identities employed to derive (1.23), it is easy to show that $\mathbf{1}_T\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L$ and $\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_T$ are given by

$$\mathbf{1}_T\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L=\frac{4\pi e^2}{\omega qmV}\sum_{ll'\mathbf{k}}\frac{\mathbf{1}_T\cdot(l\mathbf{k}|\mathbf{p}_e+\hbar\mathbf{k}|l'\mathbf{k}+\mathbf{q})(l'\mathbf{k}+\mathbf{q}|l\mathbf{k})\hat{q}[f_0(E_{l'\mathbf{k}+\mathbf{q}})-f_0(E_{l\mathbf{k}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}},$$
$$\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_T=\frac{4\pi e^2}{\omega qmV}\sum_{ll'\mathbf{k}}\frac{\hat{q}(l\mathbf{k}|l'\mathbf{k}+\mathbf{q})(l'\mathbf{k}+\mathbf{q}|\mathbf{p}_e+\hbar\mathbf{k}|l\mathbf{k})\cdot\mathbf{1}_T[f_0(E_{l'\mathbf{k}+\mathbf{q}})-f_0(E_{l\mathbf{k}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}. \quad (1.24)$$

When the limit $\mathbf{q}\to0$ is taken, the dielectric constant tensor becomes

$$\epsilon(0,\omega)=1-\frac{4\pi e^2N}{mV\omega^2}+\frac{4\pi e^2}{m^2V\omega^2}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\mathbf{p}_e|l'\mathbf{k})(l'\mathbf{k}|\mathbf{p}_e|l\mathbf{k})[f_0(E_{l'\mathbf{k}})-f_0(E_{l\mathbf{k}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}}}. \quad (1.25)$$

In a crystal of cubic symmetry, the sum over the star of $\mathbf{k}$, $\sum_{\mathbf{k}^*}(l\mathbf{k}|\mathbf{p}_e|l'\mathbf{k})(l'\mathbf{k}|\mathbf{p}_e|l\mathbf{k})$, is a multiple of the unit dyadic, and consequently $\boldsymbol{\varepsilon}(0,\omega)$ is isotropic. Thus, with the approximations made to get Eq. (1.19), as $\mathbf{q}\to0$ the longitudinal and transverse dielectric constants become equal and $\mathbf{1}_T\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_L$ and $\mathbf{1}_L\cdot\boldsymbol{\varepsilon}(\mathbf{q},\omega)\cdot\mathbf{1}_T$ vanish. This is true for an arbitrary direction of propagation in a cubic material, but will not in general hold in the case of crystals of lower symmetry.

II. LOCAL-FIELD CORRECTIONS

In this section we will develop the theory of the longitudinal dielectric constant with local-field corrections, for a cubic insulating solid, in the case of wave-

lengths large relative to the lattice constant but small relative to the over-all crystal dimensions. Local-field effects arise in a real solid because the microscopic electric field varies rapidly over the unit cell. Consequently, the macroscopic field, which is the average of the microscopic field over a region large compared with the lattice constant but small compared with the wavelength $2\pi/q$, is not in general the same as the effective or local field which polarizes the charge in the crystal. For example, suppose a slowly varying external potential

$$\phi^{\text{ext}}=\phi^{\text{ext}}(\mathbf{q},\omega)\exp i(\mathbf{q}\cdot\mathbf{r}-\omega t) \tag{2.1}$$

is applied to the crystal. The total potential $\phi=\phi^{\text{ext}}+\phi^{\text{ind}}$ will in general contain rapidly varying terms with wave vector $\mathbf{q}+\mathbf{K}$, where $\mathbf{K}$ is a vector of the reciprocal lattice:

$$\phi=\sum_{\mathbf{K}}\phi(\mathbf{q},\mathbf{K},\omega)\exp i[(\mathbf{q}+\mathbf{K})\cdot\mathbf{r}-\omega t]. \tag{2.2}$$

The potential ϕ is the microscopic potential and determines how the charge in the crystal is polarized. The macroscopic potential $\langle\phi\rangle_{\text{av}}$ is clearly given by

$$\langle\phi\rangle_{\text{av}}=\phi(\mathbf{q},0,\omega)\exp i(\mathbf{q}\cdot\mathbf{r}-\omega t), \tag{2.3}$$

since $\exp(i\mathbf{q}\cdot\mathbf{r})$ is nearly constant over the averaging region while $\exp[i(\mathbf{q}+\mathbf{K})\cdot\mathbf{r}]$, $(\mathbf{K}\neq0)$, is very rapidly varying. The derivation in Sec. I assumed a potential of form Eq. (2.1) instead of Eq. (2.2). In other words, the distinction between the microscopic and macroscopic fields and potentials was neglected, with the result that no local-field corrections were obtained. In order to obtain the longitudinal dielectric constant with local-field corrections, a total potential of the form Eq. (2.2) must be assumed (with $\mathbf{A}=0$), and the induced potential,

$$\phi^{\text{ind}}=\sum_{\mathbf{K}}\phi^{\text{ind}}(\mathbf{q},\mathbf{K},\omega)\exp i[(\mathbf{q}+\mathbf{K})\cdot\mathbf{r}-\omega t], \tag{2.4}$$

must be calculated. The longitudinal dielectric constant is obtained from the *macroscopic* total and induced potentials[4] according to an alternative form of Eq. (1.2),

$$\hat{q}\cdot\boldsymbol{\varepsilon}^L(\mathbf{q},\omega)\cdot\hat{q}=1-\langle\phi^{\text{ind}}\rangle_{\text{av}}(\mathbf{q},\omega)/\langle\phi\rangle_{\text{av}}(\mathbf{q},\omega). \tag{2.5}$$

Using Eq. (2.3), this is

$$\hat{q}\cdot\boldsymbol{\varepsilon}^L(\mathbf{q},\omega)\cdot\hat{q}=1-\phi^{\text{ind}}(\mathbf{q},0,\omega)/\phi(\mathbf{q},0,\omega). \tag{2.6}$$

The right-hand side of Eq. (2.6) is easily evaluated in a formal manner. A calculation analogous to that of Sec. I gives the relation between ϕ^{ind} and ϕ as

$$\phi^{\text{ind}}(\mathbf{q},\mathbf{K},\omega)=|\mathbf{q}+\mathbf{K}|^{-2}\sum_{\mathbf{K}'}G(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)\times\phi(\mathbf{q},\mathbf{K}',\omega), \tag{2.7}$$

with

$$G(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)=\frac{4\pi e^2}{V}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\exp(-i\mathbf{K}\cdot\mathbf{r}_e)|l'\mathbf{k}+\mathbf{q})(l'\mathbf{k}+\mathbf{q}|\exp(i\mathbf{K}'\cdot\mathbf{r}_e)|l\mathbf{k})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}. \tag{2.8}$$

As before, the variable of integration in the matrix element has been indicated by $\mathbf{r}_e$. Let us define $\epsilon(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)$ and $\epsilon^{-1}(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)$ by

$$\epsilon(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)=\delta_{\mathbf{K},\mathbf{K}'}-G(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)|\mathbf{q}+\mathbf{K}|^{-2}, \tag{2.9}$$

$$\sum_{\mathbf{K}''}\epsilon(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}'',\,\omega)\times\epsilon^{-1}(\mathbf{q}+\mathbf{K}'',\,\mathbf{q}+\mathbf{K}',\,\omega)=\delta_{\mathbf{K},\mathbf{K}'}. \tag{2.10}$$

[The quantity $\epsilon^{-1}(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)$ is just the dielectric response function of Schwinger and Martin.[5] Equations (2.9) and (2.10) have been given by Falk,[6] who treats the nearly free electron case.]

Rewrite Eq. (2.7) as

$$|\mathbf{q}+\mathbf{K}|^2[\phi(\mathbf{q},\mathbf{K},\omega)-\phi^{\text{ind}}(\mathbf{q},\mathbf{K},\omega)]=\sum_{\mathbf{K}'}\epsilon(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)|\mathbf{q}+\mathbf{K}'|^2\phi(\mathbf{q},\mathbf{K}',\omega), \tag{2.11}$$

and note that $\phi(\mathbf{q},\mathbf{K},\omega)-\phi^{\text{ind}}(\mathbf{q},\mathbf{K},\omega)=\phi^{\text{ext}}(\mathbf{q},\mathbf{K},\omega)=\phi^{\text{ext}}(\mathbf{q},\omega)\delta_{\mathbf{K},0}$, since the external potential (the potential due to charges located outside the crystal) is essentially constant over a unit cell of the crystal. Using Eq. (2.10) we find

$$\phi(\mathbf{q},\mathbf{K},\omega)=|\mathbf{q}+\mathbf{K}|^{-2}\epsilon^{-1}(\mathbf{q}+\mathbf{K},\,\mathbf{q},\,\omega)q^2\phi^{\text{ext}}(\mathbf{q},\omega), \tag{2.12}$$

giving[7]

$$\hat{q}\cdot\epsilon(\mathbf{q},\omega)\cdot\hat{q}=1/\epsilon^{-1}(\mathbf{q},\mathbf{q},\omega). \tag{2.13}$$

Thus, the problem of finding the dielectric constant with local field corrections reduces to that of solving the integral equation

$$\epsilon^{-1}(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}',\,\omega)=\delta_{\mathbf{K},\mathbf{K}'}+\sum_{\mathbf{K}''}G(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}'',\,\omega)|\mathbf{q}+\mathbf{K}''|^{-2}\times\epsilon^{-1}(\mathbf{q}+\mathbf{K}'',\,\mathbf{q}+\mathbf{K}',\,\omega), \tag{2.14}$$

obtained by combining Eqs. (2.9) and (2.10).

The main purpose of this section is to develop a systematic method of approximating the integral equation (2.14). This will be accomplished by means of two successive transformations. First, the integral equation will be transformed from the $\mathbf{K}$ representation to an $\mathbf{r}$ representation, where $\mathbf{r}$ is a continuous variable confined to a unit cell of the real lattice centered about the origin. In this representation, the kernel of the integral equation will be split into two parts, K^L and K^S. These describe the influence on a given cell of the field of the polarized charge in all other cells (K^L) and of the field of the polarized charge in the same cell (K^S), and are connected, respectively, with the local field and self-polarization corrections. A second transformation will

[4] L. Rosenfeld, *Theory of Electrons* (North-Holland Publishing Company, Amsterdam, 1951), Chap. 2.

[5] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[6] D. S. Falk, Phys. Rev. **118**, 105 (1960).

[7] This equation has been independently obtained by N. Wiser. I am indebted to M. H. Cohen and N. Wiser for communicating their results prior to publication.

then be made by expanding K^L in a multipole series, leading to an integral equation in what might be termed a multipole representation. This equation can be solved approximately by neglecting all but the first P multipole moments. The case when only dipole moments are retained will be worked out explicitly, and leads to a Lorenz-Lorentz formula modified by self-polarization corrections.

In the equations that follow, the ω dependence of ϵ^{-1} and G will no longer be indicated explicitly. To transform to the $\mathbf{r}$ representation let us define

$$\epsilon^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}')=\sum_{\mathbf{K},\mathbf{K}'} e^{i\mathbf{K}\cdot\mathbf{r}}e^{-i\mathbf{K}'\cdot\mathbf{r}'}\epsilon^{-1}(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}'),\quad(2.15)$$

$$G(\mathbf{q},\mathbf{r},\mathbf{r}')=\sum_{\mathbf{K},\mathbf{K}'} e^{i\mathbf{K}\cdot\mathbf{r}}e^{-i\mathbf{K}'\cdot\mathbf{r}'}G(\mathbf{q}+\mathbf{K},\,\mathbf{q}+\mathbf{K}').\quad(2.16)$$

The integral equation becomes

$$\begin{aligned}\epsilon^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}')&\\=v_a\sum_j\delta(\mathbf{r}-\mathbf{r}'-\mathbf{R}_j)&+\frac{1}{v_a^2}\int_0 d\mathbf{r}_1d\mathbf{r}_1'\\\times G(\mathbf{q},\mathbf{r},\mathbf{r}_1)&\left[\frac{v_a}{4\pi}\sum_j\frac{\exp i\mathbf{q}\cdot(\mathbf{R}_j+\mathbf{r}_1'-\mathbf{r}_1)}{|\mathbf{R}_j+\mathbf{r}_1'-\mathbf{r}_1|}\right]\\&\times\epsilon^{-1}(\mathbf{q},\mathbf{r}_1',\mathbf{r}'),\quad(2.17)\end{aligned}$$

where $\mathbf{R}_j$ are the vectors of the real lattice and the continuous variable $\mathbf{r}$ is confined to a unit cell centered at $\mathbf{R}_j=0$. The inverse of the dielectric constant is obtained from

$$\epsilon^{-1}(\mathbf{q},\mathbf{q})=\frac{1}{v_a^2}\int_0 d\mathbf{r}d\mathbf{r}'\,\epsilon^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}').\quad(2.18)$$

The kernel of the integral equation,

$$K(\mathbf{q},\mathbf{r},\mathbf{r}_1')=\frac{1}{v_a}\int_0 d\mathbf{r}_1\,G(\mathbf{q},\mathbf{r},\mathbf{r}_1)\frac{v_a}{4\pi}\times\sum_j\frac{\exp i\mathbf{q}\cdot(\mathbf{R}_j+\mathbf{r}_1'-\mathbf{r}_1)}{|\mathbf{R}_j+\mathbf{r}_1'-\mathbf{r}_1|},\quad(2.19)$$

can be divided into two parts,

$$K(\mathbf{q},\mathbf{r},\mathbf{r}_1')=K^L(\mathbf{q},\mathbf{r},\mathbf{r}_1')+K^S(\mathbf{q},\mathbf{r},\mathbf{r}_1').\quad(2.20)$$

Since several kernels similar in structure will be introduced in the course of the derivation, we specify them all through the functional form

$$J[\mathbf{q},f,g]=\frac{4\pi e^2}{V}\sum_{\substack{ll'\mathbf{k}\\(l\neq l')}}\frac{(l\mathbf{k}|f|l'\mathbf{k}+\mathbf{q})(l'\mathbf{k}+\mathbf{q}|g|l\mathbf{k})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}.\quad(2.21)$$

(No terms with $l=l'$ appear in the summation in the case of an insulator because all bands are either empty or full.) The kernels $G(\mathbf{q},\mathbf{r},\mathbf{r}_1')$, $K^L(\mathbf{q},\mathbf{r},\mathbf{r}_1')$, and $K^S(\mathbf{q},\mathbf{r},\mathbf{r}_1')$ are obtained from this form by the substitutions

$$\begin{aligned}G,K^L,K^S:&\quad f=V_a\textstyle\sum_j\delta(\mathbf{r}-\mathbf{r}_e-\mathbf{R}_j),\\G:&\quad g=v_a\textstyle\sum_j\delta(\mathbf{r}_1'-\mathbf{r}_e-\mathbf{R}_j),\\K^L:&\quad g=(v_a/4\pi)\exp[i\mathbf{q}\cdot(\mathbf{r}_1'-\mathbf{r}_e)]\\&\quad\times\textstyle\sum_j'\exp(i\mathbf{q}\cdot\mathbf{R}_j)/|\mathbf{r}_e-\mathbf{r}_1'-\mathbf{R}_j|,\quad(2.22)\\K^S:&\quad g=(v_a/4\pi)\exp[i\mathbf{q}\cdot(\mathbf{r}_1'-\mathbf{r}_e)]/|\mathbf{r}_e-\mathbf{r}_1'|.\end{aligned}$$

The prime on the sum defining g in K^L means that the term with $\mathbf{R}_j=0$ is to be omitted.

Since $\mathbf{r}_1'$ and $\mathbf{r}_e$ are restricted to lie within a unit cell centered about the origin, $|\mathbf{r}_e|/|\mathbf{r}_1'-\mathbf{R}_j|<1$ for all $\mathbf{R}_j\neq0$ and the multipole expansion

$$\frac{v_a}{4\pi}\sum_j{}'\frac{\exp i\mathbf{q}\cdot\mathbf{R}_j}{|\mathbf{r}_e-\mathbf{r}_1'-\mathbf{R}_j|}=\sum_{p=0}^{\infty}(\mathbf{r}_e)^p\cdot\mathbf{T}_p(\mathbf{q},\mathbf{r}_1')\quad(2.23)$$

is valid. Equation (2.23) serves as definition of the expansion coefficient $\mathbf{T}_p$. Substituting Eq. (2.17) into Eq. (2.18), splitting the kernel according to Eq. (2.20), making the multipole expansion of Eq. (2.23) and letting $\mathbf{q}\to0$ results in

$$\epsilon^{-1}(0,0)=1-i\sum_{p=1}^{\infty}\hat{q}\cdot\mathbf{K}_{1p}{}^L\cdot\mathbf{B}_p{}^L+\mathbf{B}_1{}^S\cdot\hat{q}.\quad(2.24)$$

The quantity $\mathbf{B}_p{}^L$ is defined by

$$\mathbf{B}_p{}^L=\lim_{\mathbf{q}\to0}\left[\frac{q}{v_a^2}\int_0 d\mathbf{r}'d\mathbf{r}_1'\exp(i\mathbf{q}\cdot\mathbf{r}_1')T_p(\mathbf{q},\mathbf{r}_1')\epsilon^{-1}(\mathbf{q},\mathbf{r}_1',\mathbf{r}')\right],\quad(2.25)$$

and $\mathbf{B}_1{}^S$ is the $p=1$ case of

$$\mathbf{B}_p{}^S=\lim_{\mathbf{q}\to0}\left[\frac{-iq}{v_a^2}\int_0 d\mathbf{r}'d\mathbf{r}_1'\mathbf{K}_p{}^S(\mathbf{r}_1')\epsilon^{-1}(\mathbf{q},\mathbf{r}_1',\mathbf{r}')\right].\quad(2.26)$$

The kernels $\mathbf{K}_{1p}{}^L$ and $\mathbf{K}_p{}^S(\mathbf{r}_1')$ are obtained from Eq. (2.21) by substituting

$$\begin{aligned}K_{1p}{}^L:&\quad f=\mathbf{r}_e,\ g=(\mathbf{r}_e)^p\\K_p{}^S(\mathbf{r}_1'):&\quad f=(\mathbf{r}_e)^p\exp(i\mathbf{q}\cdot\mathbf{r}_e),\quad(2.27)\\&\quad g=(v_a/4\pi)\exp[i\mathbf{q}\cdot(\mathbf{r}_1'-\mathbf{r}_e)]/|\mathbf{r}_1'-\mathbf{r}_e|.\end{aligned}$$

In deriving Eq. (2.24), the relation

$$\mathbf{q}\cdot(l'\mathbf{k}+\mathbf{q}|\mathbf{r}_e\exp(i\mathbf{q}\cdot\mathbf{r}_e)|l\mathbf{k}) = -i(l'\mathbf{k}+\mathbf{q}|l\mathbf{k})+O(q^2), \quad (2.28)$$

valid when $l\neq l'$, has been used.

The kernels $\mathbf{K}_{1p}{}^L$ and $\mathbf{K}_p{}^S(\mathbf{r}_1')$ are known quantities. In order to complete the set of equations, expressions for $\mathbf{B}_p{}^L$ and $\mathbf{B}_p{}^S$ must be derived. Multiplying Eq. (2.17) by $q\exp(i\mathbf{q}\cdot\mathbf{r}_1')\mathbf{T}_p(\mathbf{q},\mathbf{r}_1')$ and integrating gives

$$\mathbf{B}_p{}^L=i\hat{q}\delta_{p,1}+\sum_{k=1}^{\infty}(-1)^k\binom{k+p}{k}[\lim_{\mathbf{q}\to 0}\mathbf{T}_{k+p}(\mathbf{q},0)] \cdot(i\mathbf{B}_k{}^S+\sum_{n=1}^{\infty}\mathbf{K}_{kn}{}^L\cdot\mathbf{B}_n{}^L), \quad (2.29)$$

where the kernel $\mathbf{K}_{kn}{}^L$ is obtained from Eq. (2.21), by the substitution

$$\mathbf{K}_{kn}{}^L:\quad f=(\mathbf{r}_e)^k,\quad g=(\mathbf{r}_e)^n. \quad (2.30)$$

In the first term on the right-hand side of Eq. (2.29) the evaluation

$$\lim_{\mathbf{q}\to 0}\frac{q}{v_a}\int_0 d\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}}\mathbf{T}_p(\mathbf{q},\mathbf{r}) \begin{cases} =i\hat{q},\ p=1, \\ =0,\ p>1, \end{cases} \quad (2.31)$$

has been used.[8] Finally, an equation for the $\mathbf{B}_p{}^S$ must be derived. Let us write

$$D(\mathbf{q},\mathbf{r},\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}')-(1/v_a)K^S(\mathbf{q},\mathbf{r},\mathbf{r}'), \quad (2.32)$$

and define $D^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}_1)$ by

$$(1/v_a)\int_0 d\mathbf{r}_1\, D^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}_1)D(\mathbf{q},\mathbf{r}_1,\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}'). \quad (2.33)$$

Then we can write

$$\epsilon^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}') =D^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}')+(1/v_a{}^2)\int_0 d\mathbf{r}_1 d\mathbf{r}_1' D^{-1}(\mathbf{q},\mathbf{r},\mathbf{r}_1) \times K^L(\mathbf{q},\mathbf{r}_1,\mathbf{r}_1')\epsilon^{-1}(\mathbf{q},\mathbf{r}_1',\mathbf{r}'). \quad (2.34)$$

Multiplying Eq. (2.34) by $-iq\mathbf{K}_p{}^S(\mathbf{r})/v_a{}^2$, integrating and making a multipole expansion gives

$$\mathbf{B}_p{}^S=\frac{-i}{v_a{}^2}\sum_{n=1}^{\infty}\int_0 d\mathbf{r}d\mathbf{r}_1 \times\mathbf{K}_p{}^S(\mathbf{r})D^{-1}(0,\mathbf{r},\mathbf{r}_1)\mathbf{K}_n{}^L(\mathbf{r}_1)\cdot\mathbf{B}_n{}^L, \quad (2.35)$$

where $\mathbf{K}_n{}^L(\mathbf{r}_1)$ is defined by

$$\mathbf{K}_n{}^L(\mathbf{r}_1):\quad f=v_a\textstyle\sum_j\delta(\mathbf{r}_1-\mathbf{r}_e-\mathbf{R}_j),\quad g=(\mathbf{r}_e)^n. \quad (2.36)$$

[8] This is easily obtained by using

$$\frac{v_a}{4\pi}\sum_j\frac{\exp(i\mathbf{q}\cdot\mathbf{R}_j)}{|\mathbf{r}_e-\mathbf{r}-\mathbf{R}_j|}=\sum_{\mathbf{K}}\frac{\exp[i(\mathbf{q}+\mathbf{K})\cdot(\mathbf{r}_e-\mathbf{r})]}{|\mathbf{q}+\mathbf{K}|^2}.$$

It should be noted that terms involving $\mathbf{B}_0{}^L$ have been omitted from Eqs. (2.24), (2.29) and (2.35) because they vanish in the limit $\mathbf{q}\to 0$. For example, the coefficient of $\mathbf{B}_0{}^L$ in Eq. (2.24) contains a factor $(l'\mathbf{k}+\mathbf{q}|\exp(-i\mathbf{q}\cdot\mathbf{r}_e)|l\mathbf{k})$. Since

$$(l'\mathbf{k}+\mathbf{q}|\exp(-i\mathbf{q}\cdot\mathbf{r}_e)|l\mathbf{k})=(l'\mathbf{k}+\mathbf{q}|l\mathbf{k}) -i\mathbf{q}\cdot(l'\mathbf{k}+\mathbf{q}|\mathbf{r}_e|l\mathbf{k})+O(q^2)=O(q^2)$$

and

$$\mathbf{B}_0{}^L=\frac{q}{v_a}\int_0 d\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}}\frac{v_a}{4\pi}\sum_j{}'\frac{\exp(i\mathbf{q}\cdot\mathbf{R}_j)}{|\mathbf{r}+\mathbf{R}_j|} \sim\frac{q}{v_a}\int_0 d\mathbf{r}\sum_{\mathbf{K}}\frac{\exp(i\mathbf{K}\cdot\mathbf{r})}{|\mathbf{K}+\mathbf{q}|^2}=\frac{1}{q},$$

the term drops out as $\mathbf{q}\to 0$. Similarly, the terms in Eq. (2.29) and Eq. (2.35) involving $\mathbf{B}_0{}^L$ do not contribute.

The preceding manipulations have replaced the integral equation (2.14) for $\epsilon^{-1}(\mathbf{q}+\mathbf{K},\mathbf{q}+\mathbf{K}')$ by Eqs. (2.29) and (2.35), which together constitute an integral equation for $\mathbf{B}_p{}^L(p=1,\cdots,\infty)$, and the integral Eq. (2.33) for D^{-1}. Once the quantities $\mathbf{B}_p{}^L$ and $\mathbf{B}_1{}^S$ are known, the dielectric constant with local-field corrections can be calculated from Eq. (2.24). The point of this formal rearrangement is that it is now possible to make an approximation with a clear physical significance which makes Eqs. (2.29) and (2.35) easily soluble. This is simply to neglect all the $\mathbf{B}_p{}^L$ with p greater than some integer P. This means roughly that we are approximating the influence on a given cell of the charge in any other cell by the first P multipole moments of this charge. In many cases, we expect very good results to be obtained for a small value of P. The most familiar case is that of $P=1$ (dipole approximation). Utilizing the fact that the only second-order tensor compatible with cubic symmetry is the isotropic tensor, and noting that the inhomogeneous term in the equation for $\mathbf{B}_1{}^L$ is a vector parallel to $\mathbf{q}$, Eqs. (2.24), (2.29), and (2.35) become

$$\begin{aligned} &\epsilon^{-1}(0,0)=1-i\mathbf{B}_1{}^L\cdot\hat{q}\hat{q}\cdot\mathbf{K}_{11}{}^L\cdot\hat{q}+\mathbf{B}_1{}^S\cdot\hat{q}, \\ &\mathbf{B}_1{}^L\cdot\hat{q}=i-2[\lim_{\mathbf{q}\to 0}\hat{q}\cdot\mathbf{T}_2(\mathbf{q},0)\cdot\hat{q}] \times(i\mathbf{B}_1{}^S\cdot\hat{q}+\hat{q}\cdot\mathbf{K}_{11}{}^L\cdot\hat{q}\mathbf{B}_1{}^L\cdot\hat{q}), \\ &\mathbf{B}_1{}^S\cdot\hat{q}=-(i/v_a{}^2)\hat{q}\cdot\int_0 d\mathbf{r}d\mathbf{r}_1 \times\mathbf{K}_1{}^S(\mathbf{r})D^{-1}(0,\mathbf{r},\mathbf{r}_1)K_1{}^L(\mathbf{r}_1)\cdot\hat{q}\mathbf{B}_1{}^L\cdot\hat{q} \\ &\equiv -i4\pi C_1\mathbf{B}_1{}^L\cdot\hat{q}. \end{aligned} \quad (2.37)$$

It is easy to show that

$$\hat{q}\cdot\mathbf{K}_{11}{}^L\cdot\hat{q}=-4\pi\alpha, \quad (2.38)$$

where α is the polarizability calculated without making local-field corrections. [The second term on the right of Eq. (1.21) is just $4\pi\alpha$.] The dipole sum $\hat{q}\cdot\mathbf{T}_2(\mathbf{q},0)\cdot\hat{q}$ is not absolutely convergent. However, if we evaluate it in a crystal of finite diameter L, letting $L\rightarrow\infty$ and $q\rightarrow 0$ while keeping $qL\gg 1$, it has the value $-\frac{1}{3}$, independent of the crystal shape.[9] This evaluation procedure is the one that makes sense physically for wavelengths in the infrared, visible and near ultraviolet. For such wavelengths and for a typical crystal of dimension L and lattice constant a, the inequality $qL\gg 1$ holds. However, since $qa\ll 1$ and since the matrix elements and energies appearing in the kernels vary appreciably only when q changes by an amount of order $1/a$, we can still take the limit $\mathbf{q}\rightarrow 0$ in the kernels.

Using these evaluations, Eqs. (2.37) may be readily solved to yield

$$\frac{1}{\epsilon^{-1}(0,0)}=1+\frac{4\pi(\alpha-C_1)}{1-(4\pi/3)(\alpha-C_1)}. \tag{2.39}$$

This is the usual Lorenz-Lorentz formula, modified by the subtraction from α of

$$C_1=(1/4\pi v_a^2)\hat{q}\cdot\int_0 dr d\mathbf{r}_1 \times\mathbf{K}_1{}^S(\mathbf{r})\langle\mathbf{r}|[1-(1/v_a)K^S]^{-1}|\mathbf{r}_1\rangle\mathbf{K}_1{}^L(\mathbf{r}_1)\cdot\hat{q}. \tag{2.40}$$

This self-polarization correction takes just the form that is expected on the basis of a simple classical model. If we compute the dielectric constant of a macroscopic cubic lattice of uniform spheres composed of material of polarization per unit volume α, we find that the dielectric constant is determined by a Lorenz-Lorentz formula, except that α is replaced by

$$\alpha'=\alpha-\frac{4\pi\alpha^2/3}{1+4\pi\alpha/3}. \tag{2.41}$$

The subtracted term is a self-polarization correction arising from the influence on a given sphere of its surface charge. To examine the qualitative form of C_1, let us replace all the kernels K appearing in Eq. (2.40) by $-4\pi\alpha$ [cf. Eq. (2.38)]. Then we see that the correction C_1 also has the form $A\alpha^2/(1+B\alpha)$, with $A, B>0$.

With the formalism developed, higher-order corrections to the Lorenz-Lorentz formula can be obtained by taking the cutoff integer P larger than one. Note that $\mathbf{T}_p(0,0)$ is absolutely convergent for $p>2$, so no additional ambiguities regarding the method of summation appear when working to higher order. Although the calculation has been carried out for crystals of cubic symmetry in order to avoid a tensor dielectric constant, its main features would be expected to carry over to the case of arbitrary symmetry. If the restriction to insulators is dropped, intraband terms $(l=l')$ appear in Eq. (2.21). If these are treated in a free-electron approximation, which should be reasonable when motion of the conduction electrons and holes is well described by an effective mass, the generalization of the above derivation is straightforward and leads to

$$\frac{1}{\epsilon^{-1}(0,0)}=1+4\pi\alpha^{\mathrm{I}}+\frac{4\pi(\alpha^{\mathrm{II}}-C_1)}{1-(4\pi/3)(\alpha^{\mathrm{II}}-C_1)}. \tag{2.42}$$

In Eq. (2.42) α^{I} and α^{II} are, respectively, the intraband $(l=l')$ and interband $(l\neq l')$ parts of α, the polarizability without local field corrections. The restriction $qL\gg 1$, necessary to evaluate the dipole wave sum $\mathbf{T}_2(\mathbf{q},0)$, cannot be relaxed without drastically altering the derivation. In order to deal with wavelengths comparable with the macroscopic dimensions of the crystal it would be necessary to take into account surface effects, which of course has not been done in the above derivation.

ACKNOWLEDGMENT

The author wishes to thank Dr. H. Ehrenreich for suggesting the problem treated in this paper and for many helpful discussions.

APPENDIX

In order to prove that the expression for $\mathbf{j}^{\mathrm{ind}}(\mathbf{q},\omega)$ is gauge invariant, we need the auxiliary identities:

$$\textstyle\sum_l u_{l\mathbf{k}}(\mathbf{r})u_{l\mathbf{k}}{}^*(\mathbf{r}')=v_a\sum_j\delta(\mathbf{r}-\mathbf{r}'-\mathbf{R}_j), \tag{A1}$$

$$\mathbf{q}\cdot(l\mathbf{k}|\mathbf{p}_o+\hbar\mathbf{k}+\hbar\mathbf{q}/2|l'\mathbf{k}+\mathbf{q}) =(m/\hbar)(E_{l'\mathbf{k}+\mathbf{q}}-E_{l\mathbf{k}})(l\mathbf{k}|l'\mathbf{k}+\mathbf{q}). \tag{A2}$$

Expression (A1) is just the completeness relation for the periodic parts of the Bloch functions. The identity (A2) is obtained in a straightforward manner by writing $\mathbf{k}\cdot\mathbf{p}$ Schrödinger equations for $u_{l\mathbf{k}}{}^*$ and $u_{l'\mathbf{k}+\mathbf{q}}$, multiplying the former by $u_{l'\mathbf{k}+\mathbf{q}}$, the latter by $u_{l\mathbf{k}}{}^*$, and subtracting. We may also regard (A2) as the result of expanding

$$\int d\mathbf{r}\exp(-i\mathbf{q}\cdot\mathbf{r})[\nabla\cdot\langle l\mathbf{k}|\mathbf{j}_{\mathrm{op}}{}^{(0)}(\mathbf{r})|l'\mathbf{k}+\mathbf{q}\rangle+\partial\langle l\mathbf{k}|\rho_{\mathrm{op}}{}^{(0)}(\mathbf{r})|l'\mathbf{k}+\mathbf{q}\rangle/\partial t]=0, \tag{A3}$$

which shows that it is an expression of conservation of charge in the unperturbed theory.

[9] M. H. Cohen and F. Keffer, Phys. Rev. 99, 1128 (1955).

Let us now make the gauge transformation $\mathbf{A}\rightarrow\mathbf{A}+\mathbf{q}f(\mathbf{q},\omega)$, $\phi\rightarrow\phi+(\omega/c)f(\mathbf{q},\omega)$. Then

$$\Delta\mathbf{j}^{\text{ind}}/f(\mathbf{q},\omega)=-e^2N\mathbf{q}/mcV$$
$$-\frac{e^2}{m^2cV}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l'\mathbf{k}+\mathbf{q})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})](l'\mathbf{k}+\mathbf{q}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l\mathbf{k})\cdot\mathbf{q}}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}$$
$$+\frac{e^2\omega}{mcV}\sum_{ll'\mathbf{k}}\frac{(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l'\mathbf{k}+\mathbf{q})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})]}{\hbar\omega+E_{l\mathbf{k}}-E_{l'\mathbf{k}+\mathbf{q}}}.$$

Combining terms gives

$$\begin{aligned}\Delta\mathbf{j}^{\text{ind}}/f(\mathbf{q},\omega)&=(e^2/mcV)\{-N\hbar\mathbf{q}+\textstyle\sum_{ll'\mathbf{k}}(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l'\mathbf{k}+\mathbf{q})[f_0(E_{l\mathbf{k}})-f_0(E_{l'\mathbf{k}+\mathbf{q}})](l'\mathbf{k}+\mathbf{q}|l\mathbf{k})\}\\&=(e^2/mcV)[-N\hbar\mathbf{q}+2\textstyle\sum_{l\mathbf{k}}f_0(E_{l\mathbf{k}})\sum_{l'}(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l'\mathbf{k}+\mathbf{q})(l'\mathbf{k}+\mathbf{q}|l\mathbf{k})\\&=(e^2/mcV)[-N\hbar\mathbf{q}+2\textstyle\sum_{l\mathbf{k}}f_0(E_{l\mathbf{k}})(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}+\hbar\mathbf{q}/2\,|l\mathbf{k})]=0.\end{aligned}\quad\text{(A4)}$$

In the last step, the fact that $(l\mathbf{k}|\,\mathbf{p}_e+\hbar\mathbf{k}|l\mathbf{k})$ has odd parity under inversion of $\mathbf{k}$ has been used.

PHYSICAL REVIEW VOLUME 135, NUMBER 4B 24 AUGUST 1964

Tests of the Conserved Vector Current and Partially Conserved Axial-Vector Current Hypotheses in High-Energy Neutrino Reactions*

STEPHEN L. ADLER†
Princeton University, Princeton, New Jersey
(Received 8 April 1964)

The following theorem is proved: Consider the high-energy neutrino reaction $\nu+\alpha\rightarrow l+\beta$, with α a nucleon or nucleus, l a lepton (e or μ) and β a system of strongly interacting particles. Suppose that the mass of α and the invariant mass of β are not equal, and that the lepton mass is neglected. Then when the lepton emerges with its momentum parallel to that of the neutrino, the squared matrix element, averaged over lepton spin, depends only on the *divergences* of the vector and the axial-vector currents. Tests of the conserved vector current and the partially conserved axial-vector current hypotheses, based on the theorem, are proposed.

I. INTRODUCTION

THERE is a characteristic property of neutrino reactions at high energy which makes possible new tests of the conserved vector current[1] (CVC) and the partially conserved axial-vector current[2] (PCAC) hypotheses. Consider the reaction $\nu+\alpha\rightarrow l+\beta$, where α is a nucleon or nucleus, l is a muon or electron, and $\beta=\beta_1+\cdots+\beta_n$ is a system of strongly interacting particles. Let the four-momenta of ν, α, l, and β be, respectively, k_1, p_1, k_2, and p_2, and let the leptonic momentum transfer be $k=k_1-k_2=p_2-p_1$. We denote by M_α the mass of α, by m_l the mass of the lepton l, and by W the invariant mass of the system β. We take the neutrino mass to be zero.

Theorem 1. Suppose that $W\neq M_\alpha$ and that m_l is neglected. Consider the configuration in which the final lepton emerges with its momentum parallel to that of the incident neutrino. (We call this the *parallel configuration*.[3]) Then the squared matrix element for $\nu+\alpha\rightarrow l+\beta$, averaged over lepton spin, depends only on the divergences of the vector and the axial-vector currents.

Proof: The matrix element is

$$\mathfrak{M}=2^{-1/2}i\bar{u}_l\gamma_\lambda(1+\gamma_5)u_\nu\langle\beta|\mathcal{J}_\lambda{}^V+\mathcal{J}_\lambda{}^A|\alpha\rangle. \quad (1)$$

Squaring and averaging over lepton spin gives[4]

$$\langle|\mathfrak{M}|^2\rangle=\langle\beta|\mathcal{J}_\lambda{}^V+\mathcal{J}_\lambda{}^A|\alpha\rangle\langle\beta|\mathcal{J}_\sigma{}^V+\mathcal{J}_\sigma{}^A|\alpha\rangle^\star T_{\lambda\sigma}, \quad (2)$$

with

$$T_{\lambda\sigma}=k_{1\lambda}k_{2\sigma}+k_{1\sigma}k_{2\lambda}-k_1\cdot k_2\delta_{\lambda\sigma}+\epsilon_{\lambda\sigma\gamma\delta}k_{1\gamma}k_{2\delta}. \quad (3)$$

When m_l is neglected, k_1 and k_2 are null vectors. In the parallel configuration they are proportional. If $W\neq M_\alpha$, k_0 is nonzero,[5] and we may write

$$k_1=k_{10}k_0{}^{-1}k, \quad k_2=k_{20}k_0{}^{-1}k. \quad (4)$$

Thus,

$$T_{\lambda\sigma}=2k_{10}k_{20}k_0{}^{-2}k_\lambda k_\sigma. \quad (5)$$

Since $\langle\beta|\partial\mathcal{J}_\lambda/\partial x_\lambda|\alpha\rangle=-ik_\lambda\langle\beta|\mathcal{J}_\lambda|\alpha\rangle$, we find that

$$\langle|\mathfrak{M}|^2\rangle=2k_{10}k_{20}k_0{}^{-2}|\langle\beta|\partial(\mathcal{J}_\lambda{}^V+\mathcal{J}_\lambda{}^A)/\partial x_\lambda|\alpha\rangle|^2, \quad (6)$$

proving the theorem.

When $W=M_\alpha$, k_0 vanishes and the proof of the theorem breaks down. It is in fact well known that in the "elastic" weak reaction $\nu+N\rightarrow l+N$, a conserved vector current will contribute strongly in the forward direction.[6] We assume henceforth that $W\neq M_\alpha$.

II. TESTS OF CVC

Since the antisymmetric tensor term $\epsilon_{\lambda\sigma\gamma\delta}k_{1\gamma}k_{2\delta}$ vanishes under the hypotheses of Theorem 1, the characteristic parity-violating effects in weak interactions can arise only from vector-axial vector interference. Consequently, *if the vector current is conserved, and if m_l may be neglected, all parity violating effects must vanish in the parallel configuration.* This makes possible new experimental tests of the hypothesis that the vector current in $\Delta S=0$ leptonic reactions is conserved (CVC). Whereas previous tests have dealt with $\langle\beta|\mathcal{J}_\lambda{}^V|\alpha\rangle$ for $W\approx M_\alpha$ and various values of $k^2=(p_2-p_1)^2$, the new tests will study $\langle\beta|\mathcal{J}_\lambda{}^V|\alpha\rangle$ for $W\neq M_\alpha$ and $k^2\approx 0$.

Let us work in the lab frame, in which α is at rest. We assume that α is unpolarized. Then, if CVC is *false*, the two simplest types of parity violating term which may appear in the differential cross section, in the parallel configuration, are:

---

* Work supported in part by the U. S. Air Force Office of Scientific Research, Air Research and Development Command.

† National Science Foundation Predoctoral Fellow. Present address: Lyman Laboratory, Harvard University, Cambridge, Massachusetts.

[1] R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).

[2] M. L. Goldberger and S. B. Treiman, Phys. Rev. **110**, 1178 (1958); Y. Nambu, Phys. Rev. Letters **4**, 380 (1960); J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, Nuovo Cimento **17**, 757 (1960), and references listed there.

[3] When $m_l=0$, both k_1 and k_2 are null vectors. If the space components of two null vectors are parallel in one Lorentz frame, they are parallel in all Lorentz frames. Hence "parallel configuration" is an invariant concept.

[4] The four-vectors have an imaginary time component: $p=(\mathbf{p},p_4)=(\mathbf{p},ip_0)$. The quantity $p^\star$ is defined by $\mathbf{p}^\star=\mathbf{p}^*$, $p_4{}^\star=-p_4{}^*$, where * denotes complex conjugation.

[5] In the frame in which β is at rest, $k_0=(W^2-M_\alpha{}^2-k^2)/(2W)$. When $m_l=0$, k is a null vector, so if k_0 is nonvanishing in any Lorentz frame it is nonvanishing in all Lorentz frames.

[6] T. D. Lee and C. N. Yang, Phys. Rev. **126**, 2239 (1962).

(A) The vector triple-product terms

$$\mathbf{q}_i\cdot(\mathbf{q}_j\times\mathbf{q}_k)\,, \tag{7}$$

where $\mathbf{q}_i$, $\mathbf{q}_j$, and $\mathbf{q}_k$ are any three distinct momenta chosen from among the lepton momentum $\mathbf{k}_2$ and the momenta $\mathbf{q}_1,\cdots,\mathbf{q}_n$ of

$$\mathbf{q}_1,\cdots,\mathbf{q}_n \text{ of } \beta_1,\cdots,\beta_n;$$

(B) The vector·pseudovector terms

$$\mathbf{q}_i\cdot\boldsymbol{\sigma}\,, \tag{8}$$

where $\boldsymbol{\sigma}$ is the spin of a baryon in β and $\mathbf{q}_i$ is any momentum chosen from among $\mathbf{k}_2$ and $\mathbf{q}_1,\cdots,\mathbf{q}_n$.

Since, in the parallel configuration, $\mathbf{k}_1$ and $\mathbf{k}_2$ are proportional, and since $\mathbf{k}_1=\mathbf{k}_2+\mathbf{q}_1+\cdots+\mathbf{q}_n$, the system β must contain at least three particles if there are to be enough linearly independent vectors to construct a nonvanishing triple product. Consequently, the reaction with the lowest threshold which could show a triple product term is *two-pion production:*

$$\nu(\mathbf{k}_1)+\alpha(\mathbf{0})\to l(\mathbf{k}_2)+\alpha'(\mathbf{q}_1)+\pi(\mathbf{q}_2)+\pi(\mathbf{q}_3)\,. \tag{7a}$$

If CVC is valid, the laboratory differential cross section must contain no term $\mathbf{k}_2\cdot(\mathbf{q}_2\times\mathbf{q}_3)$. (There is only one linearly independent triple product in two pion production.) Note that to test CVC it is not necessary to observe the recoil nucleus α'; it is enough to know the initial neutrino direction and to observe the lepton and the two pions.

Because nucleon polarizations are hard to measure, lambda kaon production,

$$\nu(\mathbf{k}_1)+\alpha(\mathbf{0})\to l(\mathbf{k}_2)+\alpha'(\mathbf{q}_1)+\Lambda(\mathbf{q}_2)+K(\mathbf{q}_3)\,, \tag{8a}$$

is the reaction with the lowest threshold in which terms of type (B) could be detected in practice. The Λ, through its decay asymmetry, analyzes its own polarization. If CVC is valid, the terms $\boldsymbol{\sigma}_\Lambda\cdot\mathbf{k}_2$, $\boldsymbol{\sigma}_\Lambda\cdot\mathbf{q}_2$ and $\boldsymbol{\sigma}_\Lambda\cdot\mathbf{q}_3$ must not appear, while $\boldsymbol{\sigma}_\Lambda\cdot(\mathbf{k}_2\times\mathbf{q}_2)$, $\boldsymbol{\sigma}_\Lambda\cdot(\mathbf{k}_2\times\mathbf{q}_3)$, and $\boldsymbol{\sigma}_\Lambda\cdot(\mathbf{q}_2\times\mathbf{q}_3)$ are allowed. Similar tests of CVC may be constructed for reactions with thresholds higher than those of Eqs. (7a) and (8a).

The proposed tests of CVC are strictly valid only when m_l is neglected. However, we will see in the next section that the main lepton mass correction to the matrix element does not give rise to interference between a conserved vector current and the axial vector current. Consequently, the tests should be good in practice even if the lepton is a muon.

III. TESTS OF PCAC

A. Lepton Mass Neglected

Let us now accept the truth of CVC. Then, neglecting m_l, the matrix element in the parallel configuration depends only on $\langle\beta|\partial\mathcal{J}_\lambda{}^A/\partial x_\lambda|\alpha\rangle$. In an attempt to find a general explanation for the validity of the Goldberger-Treiman formula for pion decay, it has been postulated by Nambu, Gell-Mann, and others that $\mathcal{J}_\lambda{}^A$ is partially conserved.[2] We denote by PCAC the hypothesis that the covariant amplitudes contributing to $\langle\beta|\partial\mathcal{J}_\lambda{}^A/\partial x_\lambda|\alpha\rangle$ satisfy unsubtracted dispersion relations in the variable k^2 and that these dispersion relations, for $-M_\pi{}^2<k^2\lesssim M_\pi{}^2$ and for *all* values of the other invariants formed from four-momenta in α and β, are dominated by the one-pion pole. (M_π=the pion mass; we are of course considering only the case where the quantum numbers of α and β permit a one-pion pole.) Let $k_{0\beta}$ be the value of k_0 in the rest frame of β. If $k_{0\beta}{}^2/M_\pi{}^2\gg1$, the extrapolation from the physical value, $k^2\approx0$, to the pole at $k^2=-M_\pi{}^2$ will have little effect on the spinors and kinematics, and we have the covariant relation

$$\begin{aligned}&\langle\beta|\partial\mathcal{J}_\lambda{}^A/\partial x_\lambda|\alpha\rangle\\&\quad=-ik_\lambda\langle\beta|\mathcal{J}_\lambda{}^A|\alpha\rangle=(2k_0{}^{1/2})\mathcal{T}(\pi^++\alpha\to\beta)(k^2+M_\pi{}^2)^{-1}\\&\qquad\times(2k_0)^{1/2}\langle\pi^+|\partial\mathcal{T}_\lambda{}^A/\partial x_\lambda|0\rangle.\end{aligned} \tag{9}$$

Here $\mathcal{T}(\pi^++\alpha\to\beta)$ is the transition amplitude for $\pi^++\alpha\to\beta$, with the incident π^+ of energy k_0 and with the momentum of the incident π^+ parallel to $\mathbf{k}$. The Goldberger-Treiman relation, itself a consequence of PCAC, may be used to express the pion-decay matrix element in terms of g_A, the beta decay axial-vector coupling constant:

$$(2k_0)^{1/2}\langle\pi^+|\partial\mathcal{J}_\lambda{}^A/\partial x_\lambda|0\rangle=-i2^{1/2}M_Ng_Ag_r{}^{-1}M_\pi{}^2. \tag{10}$$

Numerically, $g_A\approx1.2\times10^{-5}M_N{}^{-2}$; M_N is the nucleon mass and g_r is the rationalized, renormalized pion-nucleon coupling constant ($g_r{}^2/4\pi\approx14$). Combining Eqs. (9) and (10) gives

$$\begin{aligned}&k_\lambda\langle\beta|\mathcal{J}_\lambda{}^A|\alpha\rangle\\&\quad=(2k_0)^{1/2}\mathcal{T}(\pi^++\alpha\to\beta)2^{1/2}M_Ng_Ag_r{}^{-1}M_\pi{}^2(k^2+M_\pi{}^2)^{-1}\\&\quad=(2k_0)^{1/2}\mathcal{T}(\pi^++\alpha\to\beta)2^{1/2}M_Ng_Ag_r{}^{-1}\\&\qquad\times[1-k^2(k^2+M_\pi{}^2)^{-1}].\end{aligned} \tag{11}$$

This equation may be used to express the weak-reaction cross section in the parallel configuration in terms of the cross section for $\pi^++\alpha\to\beta$. Before carrying through the details we will consider lepton mass corrections.

B. Lepton Mass Corrections

Up to this point we have neglected m_l. Now let us compute the principal lepton mass corrections. We will find that the lepton mass corrections, while not contributing significantly to terms of the form (vector) ·(axial vector) in the squared matrix element, make an

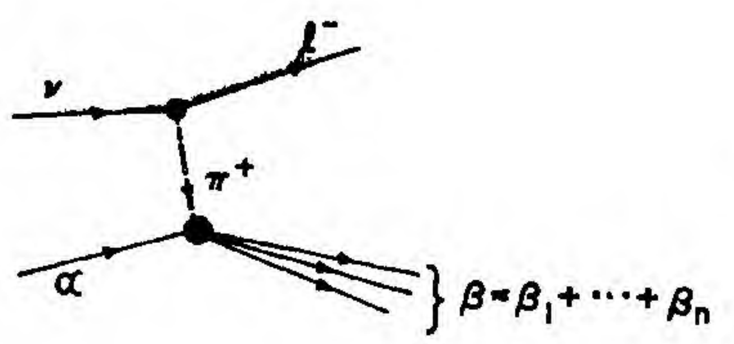


FIG. 1. Diagram giving rise to $\langle\beta|\mathcal{J}_\lambda{}^{AII}|\alpha\rangle$.

important contribution to terms of the form (axial vector)·(axial vector).

Consider the diagram shown in Fig. 1. We denote by $\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle$ the contribution of this diagram to $\langle\beta|\mathcal{J}_\lambda^{A}|\alpha\rangle$, and by $\langle\beta|\mathcal{J}_\lambda^{AI}|\alpha\rangle$ everything that is left over. It is easy to see that

$$\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle=(2k_0)^{1/2}\mathcal{T}(\pi^++\alpha\to\beta)2^{1/2}M_Ng_Ag_r^{-1}\times(-k_\lambda)(k^2+M_\pi^2)^{-1}, \quad (12)$$

from which it follows that

$$k_\lambda\langle\beta|\mathcal{J}_\lambda^{AI}|\alpha\rangle=(2k_0)^{1/2}\mathcal{T}(\pi^++\alpha\to\beta)2^{1/2}M_Ng_Ag_r^{-1}. \quad (13)$$

Although $\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle$ is a lepton mass correction, it must be retained because $m_l^2(k^2+M_\pi^2)^{-1}$ is of order unity when l is a muon. Other lepton mass corrections involve large masses in the denominator and may reasonably be neglected. Keeping terms of first order in m_l^2 in the kinematics gives[7]

$$\begin{aligned}k_1&=k_{10}k_0^{-1}k+bp_1, \quad k_2=k_{20}k_0^{-1}k+bp_1;\\ 2k\cdot p_1b&=2k_1\cdot k_2=-m_l^2k_{10}k_{20}^{-1}, \quad k^2=m_l^2k_0k_{20}^{-1}.\end{aligned} \quad (14)$$

If the vector current is conserved, it follows that

$$\begin{aligned}&\langle\beta|\mathcal{J}_\lambda^V+\mathcal{J}_\lambda^A|\alpha\rangle\langle\beta|\mathcal{J}_\sigma^V+\mathcal{J}_\sigma^A|\alpha\rangle^*T_{\lambda\sigma}\\ &=2k_{10}k_{20}k_0^{-2}|\langle\beta|k\cdot\mathcal{J}^A|\alpha\rangle|^2\\ &+2b(k_{10}+k_{20})k_0^{-1}\operatorname{Re}[\langle\beta|k\cdot\mathcal{J}^A|\alpha\rangle\langle\beta|p_1\cdot\mathcal{J}^{AII}|\alpha\rangle^*]\\ &+2b^2|\langle\beta|p_1\cdot\mathcal{J}^{AII}|\alpha\rangle|^2\\ &+\tfrac{1}{2}m_l^2k_{10}k_{20}^{-1}\{\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle^*\\ &+2\operatorname{Re}[\langle\beta|\mathcal{J}_\lambda^{AI}|\alpha\rangle\langle\beta|\mathcal{J}_\lambda^{AII}|\alpha\rangle^*]\}. \quad (15)\end{aligned}$$

We have retained m_l^2 only in terms where there is one factor $(k^2+M_\pi^2)^{-1}$ for each factor m_l^2. *No vector-axial-vector interference terms are of this form.* Substituting Eqs. (11) through (14) into Eq. (15) and performing algebraic simplification leads to the following theorem:

Theorem 2. Suppose that CVC and PCAC are true. Consider the parallel configuration in $\nu+\alpha\to l^-+\beta$, for $k_{0\beta}$ satisfying $k_{0\beta}^2/M_\pi^2\gg1$. Let m_l^2 be retained only where it occurs in the combination $m_l^2(k^2+M_\pi^2)^{-1}$. Then the invariant matrix element $\mathfrak{M}$ for $\nu+\alpha\to l^-+\beta$, squared and averaged over lepton spin, is related to the invariant matrix element[8] $\mathfrak{M}(\pi^++\alpha\to\beta)$ for $\pi^++\alpha\to\beta$, with the π^+ of energy k_0 and with the π^+ momentum parallel to $\mathbf{k}$, by

$$\langle|\mathfrak{M}|^2\rangle=\frac{4M_N^2}{g_r^2}\frac{k_{10}k_{20}}{k_0^2}g_A^2\times\left[1-\frac{m_l^2k_0}{2(M_\pi^2k_{20}+m_l^2k_0)}\right]^2|\mathfrak{M}(\pi^++\alpha\to\beta)|^2. \quad (16)$$

In computing k_{10}, k_{20}, and k_0 in Eq. (16), the lepton mass should be neglected. Then Eq. (16) will be formally covariant, since ratios of the time components of parallel null vectors, such as k_{10}/k_0 and k_{20}/k_0, are invariant quantities.

Corollary 1. Under the hypotheses of the theorem, the energy, angle, and polarization distributions of the particles in β, in the reaction $\nu+\alpha\to l^-+\beta$, will be identical with the distributions in the reaction $\pi^++\alpha\to\beta$ (for the same invariant mass W of β in the two processes).

Corollary 2. Under the hypotheses of the theorem, the lepton differential cross section $d\sigma/d\Omega_l$ of $\nu+\alpha\to l^-+\beta$, in the laboratory frame (the rest frame of α), is given by

$$\frac{d\sigma}{d\Omega_l}=\int\frac{dW}{k_0^2}\left(\frac{W}{M_\alpha}\right)^2(k_0^2-M_\pi^2)^{1/2}\left(\frac{M_\pi}{F}\right)^2\frac{k_{20}^2}{4\pi^3}g_A^2\times\left[1-\frac{m_l^2k_0}{2(M_\pi^2k_{20}+m_l^2k_0)}\right]^2\sigma(W), \quad (17)$$

where $\sigma(W)$ is the total cross section at total energy W in the β rest frame, for $\pi^++\alpha\to\beta$. Also,

$$k_0=(W^2-M_\alpha^2+M_\pi^2)/(2W),$$
$$k_{20}=(M_\alpha^2+2M_\alpha E-W^2)/(2W),$$

E is the neutrino energy in the laboratory frame, and $F=M_\pi g_r/(2M_N)\approx1.0$. The formulas for $\bar\nu+\alpha\to l^++\beta$ corresponding to those given above are obtained by replacing π^+ by π^-.

Use of Corollary 1 to test PCAC does not require knowledge of the neutrino energy spectrum, since for a given W the energy, angle and polarization distributions of the particles in β are independent of the neutrino energy E. Testing PCAC by making a quantitative comparison of $d\sigma/d\Omega_l$ with Eq. (17) of Corollary 2 does require a knowledge of the neutrino spectrum. Since all dependence on E in Eq. (17) is contained in the factors $k_{20}^2[1-\frac{1}{2}m_l^2k_0(M_\pi^2k_{20}+m_l^2k_0)^{-1}]^2$, the weighting over the spectrum is easy to carry out once the spectrum is known.

IV. EXTRAPOLATION IN k^2

Theorem 2 requires that $k_{0\beta}^2/M_\pi^2$ be much larger than unity. This condition is necessary for it to be legitimate to extrapolate from $k^2\approx m_l^2k_0k_{20}^{-1}$ to $k^2=-M_\pi^2$ in the kinematics of the reaction $\nu+\alpha\to l^-+\beta$. Since $k_{0\beta}\approx(W-M_\alpha)(W+M_\alpha)/(2W)$, the condition

[7] Actually, $k_1=k_{10}k_0^{-1}k+ak+bp_1$, $k_2=k_{20}k_0^{-1}k+ak+bp_1$, where a is of first order in m_l^2. We have dropped the term ak in Eq. (14) because it does not lead to important lepton mass corrections.

[8] The transition amplitude $\mathcal{T}(\pi^++\alpha\to\beta)$ and the invariant matrix element $\mathfrak{M}(\pi^++\alpha\to\beta)$ are related by

$$\mathcal{T}(\pi^++\alpha\to\beta)=\prod_{i,j}\left(\frac{m_i}{p_{i0}}\frac{1}{2p_{j0}}\right)^{1/2}\mathfrak{M}(\pi^++\alpha\to\beta).$$

The factor of proportionality is just the product of the normalization factors for the wave functions of π^+, α and of all the particles in β. The S matrix is given in terms of $\mathcal{T}$ by

$$S_{fi}=\delta_{fi}+(2\pi)^4i\delta(p_f-p_i)\mathcal{T}.$$

$k_{0\beta}^2/M_\pi^2 \gg 1$ will be satisfied as long as $W-M_\alpha \gtrsim 4M_\pi$. Thus, for most weak multiparticle production reactions, Theorem 2 is valid as it stands.

However, in the interesting case of single-pion production in the (3,3) resonance region, $W-M_\alpha < 4M_\pi$ and Theorem 2 must be modified. This is done by replacing $\mathfrak{M}(\pi^+ + \alpha \to \pi + \alpha')$ by $\mathfrak{M}^c(\pi^+ + \alpha \to \pi + \alpha')$, where $\mathfrak{M}^c$ is the invariant matrix element computed from the covariant amplitudes for $\pi^+ + \alpha \to \pi + \alpha'$ by using the correct kinematics, with $k^2 \approx m_l^2 k_0 k_{20}^{-1}$, for the reaction $\nu + \alpha \to l^- + \pi + \alpha'$. When α is a single nucleon N,

$$\mathfrak{M}(\pi^+ + N \to \pi + N') = (4\pi W/M_N)\chi_f^\dagger [f_1(W,\hat{q}\cdot\hat{k}) + \boldsymbol{\sigma}\cdot\hat{q}\boldsymbol{\sigma}\cdot\hat{k} f_2(W,\hat{q}\cdot\hat{k})]\chi_i, \quad (18)$$

with f_1 and f_2 the usual center-of-mass pion-nucleon scattering amplitudes,[9] and with $\hat{q}$ and $\hat{k}$ unit vectors, in the center of mass, along the momenta of the final and initial pion, respectively. Calculation of $\mathfrak{M}^c$ shows that[10]

$$\mathfrak{M}^c(\pi^+ + N \to \pi + N') = (4\pi W/M_N)\chi_f^\dagger [g_1(W,\hat{q}\cdot\hat{k}) + \boldsymbol{\sigma}\cdot\hat{q}\boldsymbol{\sigma}\cdot\hat{k} g_2(W,\hat{q}\cdot\hat{k})]\chi_i, \quad (19)$$

where

$$g_1(W,\hat{q}\cdot\hat{k}) \approx f_1(W,x),$$
$$g_2(W,\hat{q}\cdot\hat{k}) \approx [k_0/(k_0^2-M_\pi^2)^{1/2}] f_2(W,x), \quad (20)$$
$$x = [k_0/(k_0^2-M_\pi^2)^{1/2}]\hat{q}\cdot\hat{k} + [k_0(M_\pi^2+k^2)/2W(k_0^2-M_\pi^2)],$$

and where $k_0 = (W^2-M_N^2+M_\pi^2)/(2W)$. Clearly, when $k_0^2/M_\pi^2 \gg 1$ one finds that $g_{1,2}(W,\hat{q}\cdot\hat{k}) \approx f_{1,2}(W,\hat{q}\cdot\hat{k})$, as is expected.

If only the dominant (3,3) partial wave is retained, the main effect of Eq. (20) is to replace $\sigma_{3,3}(W)$ in Corollary 2 by $\sigma_{3,3}(W)k_0^2/(k_0^2-M_\pi^2)$. If, in additino, the lepton mass and nucleon recoil effects are neglected, Eq. (17) reduces to the result obtained from the static model by Bell and Berman.[11] This agreement with the static model is not surprising. The (3,3) projections of the Born terms for weak pion production and for pion-nucleon scattering can be shown to satisfy the PCAC proportionality. In the static model, the entire matrix element is determined by the (3,3) projection of the Born term and by the experimental (3,3) resonance parameters. Hence, in the static model, the weak pion production and pion-nucleon scattering matrix elements satisfy the PCAC proportionality.

Note added in proof. The considerations of this paper also apply to the decays $\Sigma^\pm \to \Lambda + e^\pm + (\nu/\bar{\nu})$, when the electron is relativistic and emerges parallel to the neutrino. For example, if CVC and PCAC are true, measurement of the differential decay rate in the parallel configuration would determine the strong $\Sigma\Lambda\Pi$ coupling constant.

ACKNOWLEDGMENTS

I wish to thank Professor S. B. Treiman for many helpful discussions and, in particular, for suggesting the most practical reactions for testing CVC.

[9] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. **106**, 1337 (1957). In Eqs. (18) through (20), isotopic spin indices have been suppressed.

[10] The calculation is done as follows: We are considering the reaction $\nu(k_1)+N(p_1) \to l(k_2)+N'(s)+\pi(q)$. Let us define the variables ν and ν_B by $\nu = -\frac{1}{2}(p_1+s)\cdot q/M_N$ and $\nu_B = \frac{1}{2}q\cdot k/M_N$. The corrected matrix element $\mathfrak{M}^c$ is given by

$$\mathfrak{M}^c = \bar{u}(s)[A^{\pi N}(\nu,\nu_B) - i\gamma\cdot q B^{\pi N}(\nu,\nu_B)]u(p_1),$$

where $A^{\pi N}(\nu,\nu_B)$ and $B^{\pi N}(\nu,\nu_B)$ are the covariant amplitudes for pion-nucleon scattering. [In pion-nucleon scattering, $\pi(q_1)+N(p_1) \to \pi(q)+N'(s)$, the variables ν and ν_B are defined by $\nu = -\frac{1}{2}(p_1+s)\cdot q/M_N$ and $\nu_B = \frac{1}{2}q\cdot q_1/M_N$.] Expressing $\mathfrak{M}^c$ in terms of pion-nucleon center of mass variables, using the kinematics appropriate to weak pion production, leads to Eq. (19).

[11] J. S. Bell and S. M. Berman, Nuovo Cimento **25**, 404 (1962). Bell and Berman neglect all lepton mass effects.

Tests of the Conserved Vector Current and Partially Conserved Axial-Vector Current Hypotheses in High-Energy Neutrino Reactions, STEPHEN L. ADLER [Phys. Rev. **135**, B963 (1964)]. The following corrections should be noted: (1) In Sec. II, line 5, m_t should be m_l; (2) the third and fourth lines after Eq. (7) should read "momenta $q_1, \cdots, q_n$ of $\beta_1, \cdots, \beta_n$"; (3) in Eq. (9), $(2k_0^{1/2})$ should be $(2k_0)^{1/2}$, and $\langle \pi^+ | \partial \mathcal{T}_\lambda^A/\partial x_\lambda | 0\rangle$ should be $\langle \pi^+ | \partial \mathcal{G}_\lambda^A/\partial x_\lambda | 0\rangle$; (4) on p. B966, column 2, line 8, "additino" should be "addition"; (5) in the *Note added in proof* $\Sigma\Lambda\Pi$ should be $\Sigma\Lambda\pi$; (6) in Eqs. (14) and (15), p_1 should be p. Delete footnote 7. (I wish to thank Dr. G. von Dardel for pointing out this correction.)

PHYSICAL REVIEW VOLUME 137, NUMBER 4B 22 FEBRUARY 1965

Consistency Conditions on the Strong Interactions Implied by a Partially Conserved Axial-Vector Current

STEPHEN L. ADLER*
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey and Bell Telephone Laboratories, Murray Hill, New Jersey

(Received 15 September 1964)

It is shown that a partially conserved $\Delta S=0$ axial-vector current ($\partial_\lambda J_\lambda{}^A=C\varphi_\pi$) implies consistency conditions involving the strong interactions alone. The most interesting of these is a relation among the symmetric isotopic-spin pion-nucleon scattering amplitude $A^{\pi N(+)}$, the pionic form factor of the nucleon $K^{NN\pi}$, and the rationalized, renormalized pion-nucleon coupling constant g_r:

$$g_r^2/M=A^{\pi N(+)}(\nu=0,\nu_B=0,\ k^2=0)/K^{NN\pi}(k^2=0).$$

[M is the nucleon mass and $-k^2$ the (mass)^2 of the initial pion. The final pion is on mass shell; the energy and momentum transfer variables ν and ν_B are defined in the text.] By using experimental pion-nucleon scattering data, we find that this relation is satisfied to within 10%. Consistency conditions involving the $\pi\pi$ and the $\pi\Lambda$ scattering amplitudes are stated.

IN 1958 Goldberger and Treiman[1] proposed a remarkable formula for the charged pion decay amplitude, which agrees with experiment to within 10%. Subsequently, Nambu, Gell-Mann and others[2] suggested that the success of the Goldberger-Treiman relation could be simply understood if it were postulated that the strangeness-conserving axial-vector current is partially conserved. The partial-conservation hypothesis leads to a number of relations connecting the weak and strong interactions, of which the Goldberger-Treiman relation is the simplest.[3] So far, only the relation for charged pion decay has been tested experimentally.

We wish to point out in this paper that, in addition to giving relations connecting the weak and strong interactions, the partially conserved axial-vector current hypothesis leads *to consistency conditions involving the strong interactions alone.*[4] This comes about, as will be explained below, because under special circumstances only the Born approximation contributes to matrix elements of the divergence of the axial-vector current. The most interesting consistency condition is a nontrivial relation among the symmetric isotopic spin pion-nucleon scattering amplitude $A^{\pi N(+)}$, the pionic form factor of the nucleon $K^{NN\pi}$, and the rationalized, renormalized pion-nucleon coupling constant g_r:

$$\frac{g_r^2}{M}=\frac{A^{\pi N(+)}(\nu=0,\ \nu_B=0,\ k^2=0)}{K^{NN\pi}(k^2=0)}. \tag{1}$$

[Here M is the nucleon mass and $-k^2$ is the (mass)^2 of the initial pion. The final pion is on mass shell; the energy and momentum transfer variables ν and ν_B are defined in Eq. (15) below.] By using experimental pion-nucleon scattering data, we find that this relation is satisfied to within 10%.

In Sec. I we define and discuss the concept of a partially conserved axial-vector current. In Sec. II, we derive the consistency condition relating the pion-nucleon scattering amplitude to the pion-nucleon coupling constant. In Sec. III, pion-nucleon dispersion relations and experimental pion-nucleon scattering data are used to test whether the consistency condition is satisfied. In Sec. IV, other consistency conditions on the strong interactions are stated.

I. DEFINITION OF PARTIALLY CONSERVED AXIAL-VECTOR CURRENT

We assume that the weak interactions between leptons and strongly interacting particles are described by a current-current effective Lagrangian of the form

$$-\mathcal{L}_{\mathrm{eff}}=J_\lambda(x)j_\lambda(x)+\text{adjoint}, \tag{2a}$$

where

$$j_\lambda(x)=(1/\sqrt{2})[\bar{\psi}_\mu\gamma_\lambda(1+\gamma_5)\psi_{\nu_\mu}+\bar{\psi}_e\gamma_\lambda(1+\gamma_5)\psi_{\nu_e}] \tag{2b}$$

is the weak current of the leptons and where J_λ is the weak current of the strongly interacting particles. Let $J_\lambda{}^V$ and $J_\lambda{}^A$ denote the vector and the axial-vector parts of the strangeness-conserving weak current

$$J_\lambda(\Delta S=0)\equiv J_\lambda{}^V+J_\lambda{}^A. \tag{2c}$$

Definition: By partially conserved axial-vector current (PCAC) we mean the hypothesis that

$$\partial_\lambda J_\lambda{}^A=-[i\sqrt{2}MM_\pi{}^2g_A(0)/g_rK^{NN\pi}(0)]\varphi_\pi+R. \tag{3}$$

Here M is the nucleon mass, M_π is the pion mass, $g_A(0)$ is the β-decay axial-vector coupling constant [$g_A(0)=1.2\cdot10^{-5}/M^2$], g_r is the rationalized, renormalized pion-nucleon coupling constant ($g_r^2/4\pi=14.6$), and φ_π

* Junior Fellow. Now at Physics Department, Harvard University, Cambridge 38, Massachusetts.

[1] M. L. Goldberger and S. B. Treiman, Phys. Rev. 109, 193 (1958).

[2] Y. Nambu, Phys. Rev. Letters 4, 380 (1960); J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, Nuovo Cimento 17, 757 (1960); M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960); J. Bernstein, M. Gell-Mann, and W. Thirring, Nuovo Cimento 16, 560 (1960).

[3] J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, Nuovo Cimento 17, 757 (1960); S. L. Adler, Phys. Rev. 135, B963 (1964).

[4] Related ideas have been discussed within the framework of a model calculation by K. Nishijima, Phys. Rev. 133, B1092 (1964).

is the renormalized field operator which creates the π^+. The quantity $K^{NN\pi}(0)$ is the pionic form factor of the nucleon evaluated at zero virtual pion mass; $K^{NN\pi}$ is normalized so that $K^{NN\pi}(-M_\pi^2)=1$. It is explained below how the constant multiplying φ_π in Eq. (3) is chosen. In order to give content to the definition, we must specify properties of the residual operator R. We suppose that for states α and β for which $\langle\beta|\varphi_\pi|\alpha\rangle\neq 0$, and for momentum transfer near the one pion pole at $-M_\pi^2$ [say, for $-M_\pi^2<(p_\beta-p_\alpha)^2<M_\pi^2$], the matrix element of R is much smaller than the matrix element of the pion operator term. In other words, we postulate that if $\langle\beta|\varphi_\pi|\alpha\rangle\neq 0$ and if $|(p_\beta-p_\alpha)^2|<M_\pi^2$, then

$$\frac{|\langle\beta|R|\alpha\rangle|}{[\sqrt{2}MM_\pi^2 g_A(0)/g_r K^{NN\pi}(0)]|\langle\beta|\varphi_\pi|\alpha\rangle|}\ll 1. \quad (4)$$

In what follows, we derive equalities which hold rigorously if the residual operator R is zero. If R is not zero, but satisfies the inequality of Eq. (4), the "equals" signs should be replaced by "approximately equals" signs. The magnitude of the squared momentum transfer $|(p_\beta-p_\alpha)^2|$ is understood to be always less than M_π^2.

It is not actually necessary to specify the constant in front of φ_π in the definition of PCAC. If we simply postulate that

$$\partial_\lambda J_\lambda^A = C\varphi_\pi, \quad (5)$$

the constant C may be determined as follows: Let us consider the matrix element of $\partial_\lambda J_\lambda^A$ between nucleon states $\langle N|\partial_\lambda J_\lambda^A|N\rangle$. Let p_2 and p_1 be, respectively, the four-momenta of the final and the initial nucleon, and let us denote by k the momentum transfer p_2-p_1. According to the usual invariance arguments, $\langle N|J_\lambda^A|N\rangle$ has the form

$$\langle N|J_\lambda^A|N\rangle=\left(\frac{M}{p_{20}}\frac{M}{p_{10}}\right)^{1/2}\bar{u}(p_2)[g_A(k^2)\gamma_\lambda\gamma_5 - f_A(k^2)\sigma_{\lambda\eta}k_\eta\gamma_5 - ih_A(k^2)k_\lambda\gamma_5]\tau^+u(p_1), \quad (6)$$

where $\tau^+=\frac{1}{2}(\tau_1+i\tau_2)$ is the isospin raising operator. From Eq. (6), we find that

$$\langle N|\partial_\lambda J_\lambda^A|N\rangle|_{k^2=0} = -ik_\lambda\langle N|J_\lambda^A|N\rangle|_{k^2=0} = 2Mg_A(0)\left(\frac{M}{p_{20}}\frac{M}{p_{10}}\right)^{1/2}\bar{u}(p_2)\gamma_5\tau^+u(p_1). \quad (7)$$

We also have

$$\langle N|C\varphi_\pi|N\rangle = \frac{C}{k^2+M_\pi^2}\langle N|(-\Box+M_\pi^2)\varphi_\pi|N\rangle = \frac{C}{k^2+M_\pi^2}\langle N|j_\pi|N\rangle = \frac{C}{k^2+M_\pi^2}ig_r\sqrt{2}K^{NN\pi}(k^2) \times\left(\frac{M}{p_{20}}\frac{M}{p_{10}}\right)^{1/2}\bar{u}(p_2)\gamma_5\tau^+u(p_1), \quad (8)$$

where $K^{NN\pi}(k^2)$ is the pionic form factor of the nucleon. From Eq. (8), we find

$$\langle N|C\varphi_\pi|N\rangle|_{k^2=0} = \frac{C}{M_\pi^2}ig_r\sqrt{2}K^{NN\pi}(0) \times\left(\frac{M}{p_{20}}\frac{M}{p_{10}}\right)^{1/2}\bar{u}(p_2)\gamma_5\tau^+u(p_1), \quad (9)$$

and comparing this with Eq. (7) gives

$$C=-i\sqrt{2}MM_\pi^2 g_A(0)/g_r K^{NN\pi}(0). \quad (10)$$

If we form the matrix element of $\partial_\lambda J_\lambda^A$ between the one pion state and the vacuum, we find that

$$(2k_0)^{1/2}\langle\pi^+|\partial_\lambda J_\lambda^A|0\rangle = -i\sqrt{2}MM_\pi^2 g_A(0)/g_r K^{NN\pi}(0), \quad (11)$$

which is the Goldberger-Treiman relation for charged pion decay. For general states β and α, such that $\langle\beta|\varphi_\pi|\alpha\rangle\neq 0$, we find that

$$\langle\beta|\partial_\lambda J_\lambda^A|\alpha\rangle = \frac{i\sqrt{2}MM_\pi^2 g_A(0)}{g_r K^{NN\pi}(0)} \times\frac{1}{k^2+M_\pi^2}(2k_0)^{1/2}\mathcal{T}(\pi^++\alpha\to\beta). \quad (12)$$

Here $\mathcal{T}(\pi^++\alpha\to\beta)$ is the transition amplitude for the strong reaction $\pi^++\alpha\to\beta$, where the (mass)^2 of the initial π^+ is $-k^2=-(p_\beta-p_\alpha)^2$. Thus, we see that PCAC leads to a whole class of relations connecting the weak and the strong interactions.

The definition of PCAC which we have given is *not* the same as the definition which would be suggested by a polology approach. This would be to define PCAC as the hypothesis that the covariant amplitudes contributing to $\langle\beta|\partial_\lambda J_\lambda^A|\alpha\rangle$ satisfy unsubtracted dispersion relations in the variable k^2, and that these dispersion relations, for $|k^2|<M_\pi^2$ and for *all* values of the other invariants formed from four-momenta in α and β, are dominated by the one pion pole. It is easy to see that if $\langle\beta|\partial_\lambda J_\lambda^A|\alpha\rangle$ depends on invariants other than k^2, the polology version of PCAC is ambiguous. Suppose that A is a covariant amplitude contributing to $\langle\beta|\partial_\lambda J_\lambda^A|\alpha\rangle$, and that A depends on two invariants, s and k^2. Then the polology version of PCAC implies that

$$A(s,k^2)=\bar{A}(s)/(k^2+M_\pi^2), \quad (13)$$

where $\bar{A}$ is the residue of A at $k^2=-M_\pi^2$. Let us now define a new variable $s'=s-ak^2$ and treat A as a function of independent variables s' and k^2. To evaluate the residue we set every *explicit* k^2 equal to $-M_\pi^2$. We then find from the polology version of PCAC that

$$A(s',k^2)\approx\frac{\bar{A}[(s-ak^2)+a(-M_\pi^2)]}{k^2+M_\pi^2}=\frac{\bar{A}[s'-aM_\pi^2]}{k^2+M_\pi^2}. \quad (14)$$

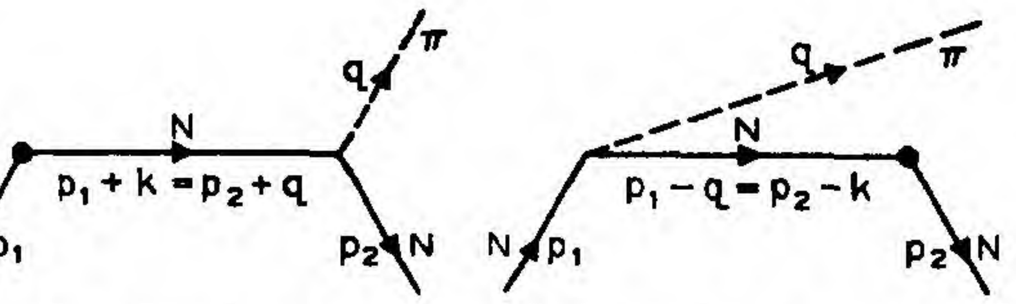


FIG. 1. Generalized Born approximation diagrams for $\langle \pi N | J_\lambda{}^A | N\rangle$. The heavy dot marks the vertex where the operator $J_\lambda{}^A$ acts.

Clearly, Eqs. (13) and (14) differ unless A has no dependence on the variable s to begin with. In other words, the polology definition of PCAC is inherently ambiguous, since the value of the residue at $k^2=-M_\pi{}^2$ depends on how the invariants other than k^2 are chosen.

This ambiguity is not present in the definition of PCAC given in Eqs. (3) and (4). The reason is that k^2 is at no point set equal to $-M_\pi{}^2$ but is kept at whatever value it has in the weak matrix element $\langle\beta|\partial_\lambda J_\lambda{}^A|\alpha\rangle$. We use the unambiguous version of PCAC in the remainder of this paper.[5]

II. CONSISTENCY CONDITION ON PION-NUCLEON SCATTERING

In the previous section we saw, in Eq. (12), that PCAC leads to relations between the strong and the weak interactions. These allow one to predict the weak interaction matrix element $\langle\beta|\partial_\lambda J_\lambda{}^A|\alpha\rangle$, if one knows the strong interaction transition amplitude $\mathcal{T}(\pi^+ + \alpha\to\beta)$. The principal point we wish to make in this paper is that there are cases in which only the Born approximation contributes to a covariant amplitude of $\langle\beta|\partial_\lambda J_\lambda{}^A|\alpha\rangle$, for appropriately chosen values of the energy, momentum transfer and other invariants on which the covariant amplitude depends. The Born approximation, in turn, is known in terms of weak and strong interaction coupling constants. Using PCAC to eliminate the weak interaction coupling constants leaves a consistency condition involving the strong interactions alone. In this section, we study the matrix element $\langle\pi N|\partial_\lambda J_\lambda{}^A|N\rangle$ and derive the consistency condition stated in Eq. (1). In Sec. IV, we discuss conditions obtained from other matrix elements of $\partial_\lambda J_\lambda{}^A$.

We begin by writing down the structure of the matrix element $\langle\pi N|J_\lambda{}^A|N\rangle$. Let p_1, p_2, and q be, respectively, the four-momenta of the initial nucleon, the final nucleon, and the final pion. The momentum transfer k is given by $k=p_2+q-p_1$. We define invariants ν and ν_B by

$$\begin{aligned}\nu&=-(p_1+p_2)\cdot k/(2M)\,,\\ \nu_B&=q\cdot k/(2M)\,.\end{aligned}\tag{15}$$

The matrix element can be decomposed into eight covariant amplitudes $A_j(\nu,\nu_B,k^2)$ according to

$$\left(\frac{p_{10}}{M}\frac{p_{20}}{M}2k_0\right)^{1/2}\langle\pi N|J_\lambda{}^A|N\rangle = \bar u(p_2)i\sum_{j=1}^{8}O_j{}^\lambda A_j(\nu,\nu_B,k^2)u(p_1)\,.\tag{16}$$

The quantities $O_j{}^\lambda$ are given by[6]

$$\begin{aligned}&O_1{}^\lambda=\tfrac{1}{2}(q\gamma_\lambda-\gamma_\lambda q)\,, && O_5{}^\lambda=ik(p_1+p_2)_\lambda\,,\\ &O_2{}^\lambda=(p_1+p_2)_\lambda\,, && O_6{}^\lambda=ikq_\lambda\,,\\ &O_3{}^\lambda=q_\lambda\,, && O_7{}^\lambda=k_\lambda\,,\\ &O_4{}^\lambda=iM\gamma_\lambda\,, && O_8{}^\lambda=ikk_\lambda\,.\end{aligned}\tag{17}$$

The amplitudes $A_j(\nu,\nu_B,k^2)$ have been chosen so that they have no kinematic singularities.[7]

The isotopic spin structure of the amplitudes $A_j(\nu,\nu_B,k^2)$ is specified by writing

$$\begin{aligned}A_j(\nu,\nu_B,k^2)&=\chi_f{}^*\psi_\alpha{}^*A_j(\nu,\nu_B,k^2)_{\alpha\beta}\psi_\beta{}^+\chi_i\,,\\ A_j(\nu,\nu_B,k^2)_{\alpha\beta}&=A_j{}^{(+)}(\nu,\nu_B,k^2)\delta_{\alpha\beta}\\ &\quad+A_j{}^{(-)}(\nu,\nu_B,k^2)\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\,.\end{aligned}\tag{18}$$

Here χ_i and χ_f are, respectively, the isospinors of the initial and final nucleon and ψ_α is the isotopic spin wave function of the final pion. [If the final pion is a $\pi^\pm$, $\psi_\alpha=2^{-1/2}(1,\pm i,0)_\alpha$, while if it is a π^0, $\psi_\alpha=(0,0,1)_\alpha$.] The quantity $\psi_\beta{}^+$ is defined by $\psi_\beta{}^+=\frac{1}{2}(1,i,0)_\beta$, so that $\psi_\beta{}^+\tau_\beta=\tau^+$. The presence of $\psi_\beta{}^+$ is just a reflection of the fact that the weak current $J_\lambda{}^A$ transforms like I_1+iI_2 under isotopic spin rotations.

Let us split each amplitude $A_j(\nu,\nu_B,k^2)_{\alpha\beta}$ into two parts,

$$A_j(\nu,\nu_B,k^2)_{\alpha\beta}=A_j{}^P(\nu,\nu_B,k^2)_{\alpha\beta}+\bar A_j(\nu,\nu_B,k^2)_{\alpha\beta}\,.\tag{19}$$

The part $A_j{}^P$ is defined as the sum of all pole terms contributing to A_j, while $\bar A_j$ is simply everything that is left over when the pole terms are removed from A_j. The amplitudes $A_j{}^P$ are calculated from the generalized Born approximation diagrams shown in Fig. 1. In each diagram, the heavy dot marks the vertex where the operator $J_\lambda{}^A$ acts. The nucleon vertex of $J_\lambda{}^A$ is given by

$$\tau^+[g_A(k^2)\gamma_\lambda\gamma_5-f_A(k^2)\sigma_{\lambda\eta}k_\eta\gamma_5-ih_A(k^2)k_\lambda\gamma_5]\,.\tag{20}$$

Evaluation of the Born diagrams gives

$$\begin{aligned}&\bar u(p_2)i\sum_{j=1}^{8}O_j{}^\lambda\chi_f{}^*\psi_\alpha{}^*A_{j\alpha\beta}{}^P\psi_\beta{}^+\chi_i u(p_1)\\ &=\bar u(p_2)\chi_f{}^*\psi_\alpha{}^*\{i\tau_\alpha\gamma_5 g_r[1/(p_2+q-iM)]\\ &\quad\times\tau^+[g_A(k^2)\gamma_\lambda\gamma_5-f_A(k^2)\sigma_{\lambda\eta}k_\eta\gamma_5-ih_A(k^2)k_\lambda\gamma_5]\\ &\quad+\tau^+[g_A(k^2)\gamma_\lambda\gamma_5-f_A(k^2)\sigma_{\lambda\eta}k_\eta\gamma_5-ih_A(k^2)k_\lambda\gamma_5]\\ &\quad\times[1/(p_1-q-iM)]i\tau_\alpha\gamma_5 g_r\}\chi_i u(p_1)\,,\end{aligned}\tag{21}$$

from which the $A_j{}^P$ are easily obtained. Since the divergence of the terms proportional to $f_A(k^2)$ vanishes

[5] In a previous paper [S. L. Adler, Phys. Rev. 135, B963 (1964)] we used the polology version of PCAC. If, instead, the definition of Eqs. (3) and (4) had been used, $\mathfrak{M}(\pi^++\alpha\to\beta)$ in Theorem 2 of the paper would simply have been the invariant matrix element for $\pi^++\alpha\to\beta$, with the initial π^+ of (mass)$^2=-k^2$.

[6] The kinematic structure of the matrix element $\langle\pi N|J_\lambda{}^A|N\rangle$ has been discussed by N. Dombey, Phys. Rev. 127, 653 (1962) and by P. Dennery, Phys. Rev. 127, 664 (1962).

[7] A simple modification of the argument used by Ball [J. S. Ball, Phys. Rev. 124, 2014 (1961)] can be used to show that the amplitudes A_j have no kinematical singularities.

identically and since the divergence of the terms proportional to $h_A(k^2)$ vanishes when $k^2=0$, we write down only the pole contributions proportional to $g_A(k^2)$:

$$A_1{}^P=\frac{g_r g_A(k^2)}{2M}\bigg[\delta_{\alpha\beta}\bigg(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\bigg)+\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\bigg(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\bigg)\bigg],$$

$$A_3{}^P=\frac{g_r g_A(k^2)}{2M}\bigg[\delta_{\alpha\beta}\bigg(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\bigg)+\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\bigg(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\bigg)\bigg]. \tag{22}$$

The amplitudes A_2 and A_4, $\cdots$, A_8 have no pole contributions proportional to $g_A(k^2)$.

Let us now evaluate

$$\langle\pi N|\partial_\lambda J_\lambda{}^A|N\rangle=-ik_\lambda\langle\pi N|J_\lambda{}^A|N\rangle$$

at $k^2=0$. Using the decomposition of $\langle\pi N|J_\lambda{}^A|N\rangle$ into covariants A_j, splitting each A_j into parts $A_j{}^P$ and $\bar{A}_j$, and evaluating the $A_j{}^P$ from Eq. (22), leads to the result that

$$[(p_{10}/M)(p_{20}/M)2k_0]^{1/2}\langle\pi N|\partial_\lambda J_\lambda{}^A|N\rangle|_{k^2=0}=\bar{u}(p_2)\chi_f{}^*\psi_\alpha{}^*M_{\alpha\beta}(\sqrt{2}\psi_\beta{}^+)\chi_i u(p_1), \tag{23}$$

with

$$M_{\alpha\beta}=A(\nu,\nu_B)_{\alpha\beta}-ikB(\nu,\nu_B)_{\alpha\beta},$$

$$A(\nu,\nu_B)_{\alpha\beta}=\frac{1}{\sqrt{2}}\{-2M\nu(\bar{A}_1+\bar{A}_2)_{\alpha\beta}+2M\nu_B\bar{A}_{3\alpha\beta}+2g_r g_A(0)\delta_{\alpha\beta}\},$$

$$B(\nu,\nu_B)_{\alpha\beta}=\frac{1}{\sqrt{2}}\bigg\{2M\bar{A}_{1\alpha\beta}-M\bar{A}_{4\alpha\beta}+2M\nu\bar{A}_{5\alpha\beta}-2M\nu_B\bar{A}_{6\alpha\beta}+g_r g_A(0)\times\bigg[\delta_{\alpha\beta}\bigg(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\bigg)+\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\bigg(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\bigg)\bigg]\bigg\}. \tag{24}$$

According to the PCAC hypothesis, we can also evaluate $\langle\pi N|\partial_\lambda J_\lambda{}^A|N\rangle$ as $\langle\pi N|C\varphi_\pi|N\rangle$. This gives

$$M_{\alpha\beta}=\frac{\sqrt{2}Mg_A(0)}{g_r K^{NN\pi}(0)}[A^{\pi N}(\nu,\nu_B,k^2=0)_{\alpha\beta}-ikB^{\pi N}(\nu,\nu_B,k^2=0)_{\alpha\beta}]$$
$$=\frac{\sqrt{2}Mg_A(0)}{g_r K^{NN\pi}(0)}\bigg\{A^{\pi N}(\nu,\nu_B,k^2=0)_{\alpha\beta}-ik\bar{B}^{\pi N}(\nu,\nu_B,k^2=0)_{\alpha\beta}-ik\frac{g_r{}^2}{2M}K^{NN\pi}(0)\times\bigg[\delta_{\alpha\beta}\bigg(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\bigg)+\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\bigg(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\bigg)\bigg]\bigg\}. \tag{25}$$

The amplitudes $A^{\pi N}(\nu,\nu_B,k^2=0)$ and $B^{\pi N}(\nu,\nu_B,k^2=0)$ describe pion-nucleon scattering with the initial pion a virtual pion of $(\text{mass})^2=-k^2=0$ and with the final pion a real pion of $(\text{mass})^2=M_\pi{}^2$.[8] We have separated off the pole terms of B (A has no pole terms); $\bar{B}$ denotes everything which is left over after this separation is made.

Comparing Eqs. (24) and (25), we see that the pole terms proportional to

$$\delta_{\alpha\beta}\bigg(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\bigg)+\tfrac{1}{2}[\tau_\alpha,\tau_\beta]\bigg(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\bigg) \tag{26}$$

are identical. This is consistent with the requirements of PCAC. A remarkable fact emerges when we consider the equation for the A amplitudes,

$$(1/\sqrt{2})[-2M\nu(\bar{A}_1+\bar{A}_2)_{\alpha\beta}+2M\nu_B\bar{A}_{3\alpha\beta}+2g_r g_A(0)\delta_{\alpha\beta}]=[\sqrt{2}Mg_A(0)/g_r K^{NN\pi}(0)]A^{\pi N}(\nu,\nu_B,k^2=0)_{\alpha\beta}. \tag{27}$$

Let us set $\nu=\nu_B=0$. Since the $\bar{A}_j$ have all pole terms removed, and since they have no kinematic singularities,

$$\lim_{\nu\to0}\nu(\bar{A}_1+\bar{A}_2)=\lim_{\nu_B\to0}\nu_B\bar{A}_3=0. \tag{28}$$

Hence at $\nu=\nu_B=k^2=0$, all the unknown amplitudes drop out. Equation (27) then becomes

$$\delta_{\alpha\beta}\frac{g_r{}^2}{M}=\frac{A^{\pi N}(\nu=0,\nu_B=0,k^2=0)_{\alpha\beta}}{K^{NN\pi}(0)}. \tag{29}$$

Decomposing $A_{\alpha\beta}{}^{\pi N}$ into symmetric and antisymmetric isotopic spin parts,

$$A_{\alpha\beta}{}^{\pi N}=A^{\pi N(+)}\delta_{\alpha\beta}+A^{\pi N(-)}\tfrac{1}{2}[\tau_\alpha,\tau_\beta] \tag{30}$$

gives

$$\frac{g_r{}^2}{M}=\frac{A^{\pi N(+)}(\nu=0,\nu_B=0,k^2=0)}{K^{NN\pi}(0)}, \tag{31}$$

$$0=A^{\pi N(-)}(\nu=0,\nu_B=0,k^2=0). \tag{32}$$

Equation (32) is automatically satisfied by virtue of the odd crossing symmetry of $A^{\pi N(-)}$. Equation (31) is a nontrivial consistency condition which must be satisfied if PCAC is true.

We saw above that the pole terms, which are the only pion-nucleon scattering terms of second order in the coupling constant g_r, do not contribute to the amplitude $A^{\pi N(+)}$. The leading term in the perturbation series for $K^{NN\pi}$ is 1. Consequently, if $A^{\pi N(+)}/K^{NN\pi}$ is expanded in a renormalized perturbation series, no term of order $g_r{}^2$ will be present. Thus it is clear that the consistency condition is not an identity in the coupling constant. This makes it fundamentally different from relations obtained from unitarity or from crossing symmetry, which are always true order by order in perturbation theory.

[8] Pion-nucleon scattering with the initial pion virtual has been discussed by E. Ferrari and F. Selleri, Nuovo Cimento **21**, 1028 (1961) and by J. Iizuka and A. Klein, Progr. Theoret. Phys. (Kyoto) **25**, 1017 (1961).

Note that a similar consistency condition cannot be derived for the B amplitudes, since the presence of the terms $2M\bar{A}_1 - M\bar{A}_4$ in Eq. (24) prevents the elimination of the unknown amplitudes $\bar{A}_1$ and $\bar{A}_4$.

In the next section, the condition of Eq. (31) is compared with experiment. Before going on to do this, let us summarize the properties of $J_\lambda{}^A$ that were actually used in the derivation. Nowhere did we use the fact that $J_\lambda{}^A$ is the weak axial-vector current which couples to the leptons. Clearly, the consistency condition may be derived if the following two conditions are met:

(i) There exists a local axial-vector current J_λ, the divergence of which is proportional to the pion field,

$$\partial_\lambda J_\lambda = C\varphi_\pi; \tag{33}$$

(ii) In the nucleon vertex of J_λ, which apart from isospin is

$$\langle N|J_\lambda|N\rangle = \bar{u}(p_2)[G(k^2)\gamma_\lambda\gamma_5 - F(k^2)\sigma_{\lambda\eta}k_\eta\gamma_5 - iH(k^2)k_\lambda\gamma_5]u(p_1), \tag{34}$$

the form factors G, F, and H are finite at $k^2=0$, and furthermore, $G(0)$ is nonvanishing. In the matrix element $\langle\pi N|J_\lambda|N\rangle$, the covariant amplitudes $A_{1,\dots,8}(\nu,\nu_B,k^2)$ are finite at $\nu=\nu_B=k^2=0$ once the poles which arise from the Born-approximation (one-particle intermediate state) diagrams are subtracted off. [Except for the requirement that $G(0)$ be nonvanishing, these conditions are necessarily satisfied if the form factors and covariant amplitudes in the two matrix elements of J_λ satisfy the usual spectral conditions, that is, if their singularities as functions of the complex variables k^2, ν and ν_B arise only from allowed intermediate states.]

Condition (ii) and the requirement of locality are essential for the derivation to go through. They are very restrictive conditions, and it is easy to find axial-vector currents which do not satisfy them but which obey Eq. (33). For instance, the current $J_\lambda{}'$ defined by

$$J_\lambda{}' = C\partial_\lambda \int D(x-x')\varphi_\pi(x')d^4x',$$
$$D(x) = -\int \frac{e^{ik\cdot x}}{(2\pi)^4 k^2} d^4k, \tag{35}$$

satisfies $\partial_\lambda J_\lambda{}' = C\varphi_\pi$, by construction. But $J_\lambda{}'$ is not local, and in the nucleon vertex of $J_\lambda{}'$, $G(k^2)\equiv 0$ and $H(k^2)\alpha 1/k^2$, so that (ii) is violated.

III. DISPERSION RELATIONS TEST OF CONSISTENCY CONDITION

In this section, we use pion-nucleon dispersion relations and experimental pion-nucleon scattering data to test whether Eq. (31) is satisfied in nature. By using dispersion relations, the on-mass-shell amplitude $A^{\pi N(+)}(\nu=0, \nu_B=0, k^2=-M_\pi{}^2)$ may be calculated from scattering data. However, Eq. (31) involves the off-mass-shell combination $A^{\pi N(+)}(\nu=0, \nu_B=0, k^2=0)/K^{NN\pi}(k^2=0)$, requiring us to use a model to calculate the difference,

$$\frac{A^{\pi N(+)}(\nu=0, \nu_B=0, k^2=0)}{K^{NN\pi}(k^2=0)} - A^{\pi N(+)}(\nu=0, \nu_B=0, k^2=-M_\pi{}^2). \tag{36}$$

We first give several alternative ways of using pion-nucleon dispersion relations to calculate the on-mass-shell amplitude. We then discuss a model for going off mass shell in k^2, and summarize the final results. In the remainder of this section, we take the charged pion mass to be unity. In these units the nucleon mass is $M=6.72$ and[9]

$$g_r{}^2/M = 27.4\pm 0.7. \tag{37}$$

The equations used in making the calculations described in this section are derived in the Appendix.

A. Evaluation of $A^{\pi N(+)}$ ($\nu=0$, $\nu_B=0$, $k^2=-1$)

We wish to evaluate the on-mass-shell amplitude $A^{\pi N(+)}(0,0\ -1)$. Since the point $\nu=\nu_B=0$ is not a physical one, we must use pion-nucleon dispersion relations to compute $A^{\pi N(+)}(0,0,-1)$ from scattering data. The fixed momentum transfer dispersion relation satisfied by $A^{\pi N(+)}(\nu,\nu_B,-1)$ is[10]

$$A^{\pi N(+)}(\nu,\nu_B,-1) = \frac{1}{\pi}\int_{\nu_0+\nu_B}^{\infty} d\nu' \,\mathrm{Im}A^{\pi N(+)}(\nu',\nu_B,-1) \times\left[\frac{1}{\nu'-\nu}+\frac{1}{\nu'+\nu}\right], \quad \nu_0 = 1+1/(2M). \tag{38}$$

Since the integral in Eq. (38) probably does not converge, it is necessary to introduce a subtraction.

1. *Threshold Subtraction*

The usual procedure is to make a subtraction at threshold. This gives

$$A^{\pi N(+)}(0,0,-1) = A^{\pi N(+)}(\nu_0,0,-1) - \frac{2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu'}{\nu'} \times \frac{\mathrm{Im}A^{\pi N(+)}(\nu',0,-1)\nu_0{}^2}{(\nu'^2-\nu_0{}^2)}, \tag{39}$$

which has a strongly convergent integral. The integrand can be calculated in terms of phase shifts. The integral

[9] The coupling constant $g_r{}^2$ is related to the coupling constant f^2 by $g_r{}^2=4\pi\cdot 4M^2f^2$. We use the value $f^2=0.081\pm0.002$ quoted by W. S. Woolcock, *Proceedings of the Aix-en-Provence International Conference on Elementary Particles* (Centre d'Etudes Nucléaires de Saclay, Seine et Oise, 1961), Vol. I, p. 459.

[10] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. **106**, 1337 (1957).

was evaluated using the phase shift analysis of Roper[11] up to a pion laboratory kinetic energy of $T_\pi=700$ MeV, where the integral was truncated. A convergence check indicated that the truncation error is small. The result is

$$\frac{2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu'}{\nu'}\frac{\mathrm{Im}A^{\pi N(+)}(\nu',0,-1)\nu_0^2}{\nu'^2-\nu_0^2}=7.4. \quad (40)$$

We make no error estimate here since Roper gives no error estimate for his phase shifts.

The threshold subtraction constant can be expressed in terms of scattering lengths by

$$\frac{A^{\pi N(+)}(\nu_0,0,-1)}{4\pi}=\left(1+\frac{1}{2M}\right)\sum_{l=0}^{\infty}\left[\tfrac{2}{3}a_{l+}^{(3/2)}+\tfrac{1}{3}a_{l+}^{(1/2)}\right]\frac{(l+1)[2(l+1)]!}{2^{l+1}[(l+1)!]^2}-2M\sum_{l=1}^{\infty}\left\{\tfrac{2}{3}\left[a_{l-}^{(3/2)}-a_{l+}^{(3/2)}\right]+\tfrac{1}{3}\left[a_{l-}^{(1/2)}-a_{l+}^{(1/2)}\right]\right\}\frac{l(2l)!}{2^l(l!)^2}, \quad (41)$$

where $a_{l\pm}^{(I)}$ is the scattering length in the channel with isospin I, orbital angular momentum l, and total angular momentum $J=l\pm\frac{1}{2}$. Equation (41) is rapidly convergent and it suffices to keep only the S-, P-, D-, and F-wave scattering lengths. Using the S- and P-wave scattering lengths quoted by Woolcock[12] and obtaining D- and F-wave scattering lengths from Roper's polynomial and resonance fits to the phase shifts, gives

$$\begin{aligned}A^{\pi N(+)}(\nu_0,0,-1)&=37.3\pm0.7,\\ A^{\pi N(+)}(0,0,-1)&=29.9\pm0.7.\end{aligned} \quad (42)$$

The threshold subtraction constant arises almost entirely from the P-wave scattering lengths. The error estimates take into account only the errors in the S- and P-wave scattering lengths quoted by Woolcock.

Alternatively, we can obtain all scattering lengths from the threshold behavior of Roper's fits to the phase shifts, giving

$$\begin{aligned}A^{\pi N(+)}(\nu_0,0,-1)&=40.7,\\ A^{\pi N(+)}(0,0,-1)&=33.3.\end{aligned} \quad (43)$$

2. *Broad Area Subtraction Method*

There is a fairly large discrepancy between Woolcock's scattering lengths and the threshold behavior of Roper's phase shifts. This suggests that it would be desirable to perform the subtraction in a manner which does not weight threshold behavior so heavily. We give a method which effectively smears the subtraction over a finite segment of the real axis and has the additional advantage of containing a built-in consistency check on the phase shift data used. Let us consider the function

$$F(\nu)=\frac{A^{\pi N(+)}(\nu,\nu_B=0,k^2=-1)}{[(\nu-\nu_0)(\nu+\nu_0)(\nu-\nu_m)(\nu+\nu_m)]^{1/2}}, \quad (44)$$

where $\nu_m>\nu_0$ lies on the physical cut. Since $F(\nu)$ approaches zero at $\nu=\infty$, we can write an unsubtracted dispersion relation

$$F(\nu)=\frac{1}{\pi}\int_{\nu_0}^{\infty}d\nu'\frac{\Delta F(\nu')}{2i}\left(\frac{1}{\nu'-\nu}+\frac{1}{\nu'+\nu}\right), \quad (45)$$

where $\Delta F(\nu')=F(\nu'+i\epsilon)-F(\nu'-i\epsilon)$ is the discontinuity of F across the cut from ν_0 to ∞. The square root in the denominator has opposite signs on the opposite sides of its cut from ν_0 to ν_m and has no cut from ν_m to ∞. Consequently,

$$\begin{aligned}\frac{\Delta F(\nu')}{2i}&=-\frac{\mathrm{Re}A^{\pi N(+)}(\nu',0,-1)}{[(\nu'-\nu_0)(\nu'+\nu_0)(\nu_m-\nu')(\nu_m+\nu')]^{1/2}}, & \nu_0<\nu'<\nu_m,\\ \frac{\Delta F(\nu')}{2i}&=\frac{\mathrm{Im}A^{\pi N(+)}(\nu',0,-1)}{[(\nu'-\nu_0)(\nu'+\nu_0)(\nu'-\nu_m)(\nu'+\nu_m)]^{1/2}}, & \nu_m<\nu'<\infty,\end{aligned} \quad (46)$$

giving

$$A^{\pi N(+)}(0,0,-1)=\frac{2}{\pi}\int_{\nu_0}^{\nu_m}\frac{d\nu'}{\nu'}\frac{\mathrm{Re}A^{\pi N(+)}(\nu',0,-1)\nu_0\nu_m}{[(\nu'-\nu_0)(\nu'+\nu_0)(\nu_m-\nu')(\nu_m+\nu')]^{1/2}}-\frac{2}{\pi}\int_{\nu_m}^{\infty}\frac{d\nu'}{\nu'}\frac{\mathrm{Im}A^{\pi N(+)}(\nu',0,-1)\nu_0\nu_m}{[(\nu'-\nu_0)(\nu'+\nu_0)(\nu'-\nu_m)(\nu'+\nu_m)]^{1/2}}. \quad (47)$$

This equation involves $\mathrm{Re}A^{\pi N(+)}$ over a segment of finite length, not just at threshold. In the limit as ν_m approaches ν_0, Eq. (47) becomes identical with Eq. (39) for the threshold subtraction. The fact that Eq. (47) involves no principal value integrals makes numerical evaluation easy.

If the exact values of $\mathrm{Re}A^{\pi N(+)}$ and $\mathrm{Im}A^{\pi N(+)}$ were used to evaluate the integrals, Eq. (47) would clearly give the same answer for all values of ν_m between ν_0 and ∞. Thus, by varying ν_m we can check the consistency of the phase shifts used to evaluate $A^{\pi N(+)}(\nu',0,-1)$.

Using the phase shift data of Roper and integrating up to a pion laboratory kinetic energy of 700 MeV gives

[11] L. D. Roper, Phys. Rev. Letters 12, 340 (1964) and private communication. We use Roper's $l_m=3$ phase shift fit for pion laboratory kinetic energy T_π in the range $0\le T_\pi\le 700$ MeV. In terms of ν and ν_B, $T_\pi=\nu-\nu_B-\nu_0$.

[12] W. S. Woolcock, Ref. 9. Woolcock's results are quoted in J. Hamilton, P. Menotti, G. C. Oades, and L. L. J. Vick, Phys. Rev. 128, 1881 (1962). The slightly different scattering lengths proposed by Hamilton *et al.* give the same result for $A^{\pi N(+)}(\nu_0,0,-1)$ as do Woolcock's.

TABLE I. $A^{\pi N(+)}$ versus length of the square-root cut, as calculated by using fixed momentum transfer dispersion relations. The upper end of the square-root cut lies at pion-nucleon center-of-mass energy $W_m=M+\omega_m$. At threshold, $\omega_m=1$, and at the peak of the (3,3) resonance, $\omega_m=2.1$. In terms of ω_m, ν_m is given by $\nu_m=\nu_B+\omega_m+\omega_m^2/2M$.

ω_m	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5
$A^{\pi N(+)}(0,0,-1)$ $(\nu_B=0)$	29.70	29.43	29.16	28.90	28.66	28.44	28.23	28.02	27.83
$A^{\pi N(+)}(0,-1/2M,-1)$ $(\nu_B=-1/2M)$	26.73	26.54	26.36	26.17	26.00	25.83	25.67	25.51	25.36
D	2.97	2.89	2.80	2.73	2.66	2.61	2.56	2.51	2.47

the results shown in Table I. It is convenient to introduce a parameter ω_m, such that the upper end of the square-root cut lies at pion-nucleon center-of-mass energy $W_m=M+\omega_m$. In terms of ω_m, the parameter ν_m is given by $\nu_m=\nu_B+\omega_m+\omega_m^2/2M$. In changing ω_m from 1.7 to 2.5, we move the upper end of the square-root cut across the peak of the (3,3) resonance, thus considerably altering the distribution of the integral between the two terms in Eq. (47). Still, the total varies by less than 10%, indicating that Roper's phase shifts are reasonably consistent with dispersion relations in the (3,3) resonance region. The end of the cut was not taken greater than $\omega_m=2.5$ to avoid introducing a large truncation error from extending the integrals only to 700 MeV. A convergence check indicated that in all cases shown in Table I the truncation error is small.

The result of this analysis may be stated as

$$A^{\pi N(+)}(0,0,-1)=28.7\pm0.9, \qquad (48)$$

where we have taken as the error estimate the variation of $A^{\pi N(+)}$ as ω_m is moved across the peak of the (3,3) resonance.

3. *Alternative Broad-Area Subtraction Method*

As a further check, we have used an alternative method to evaluate $A^{\pi N(+)}(0,0,-1)$. Let us write

$$\begin{aligned} A^{\pi N(+)}(0,0,-1)&=D+A^{\pi N(+)}(\nu=0,\nu_B=-1/2M,-1),\\ D&=A^{\pi N(+)}(\nu=0,\nu_B=0,-1) \qquad (49)\\ &\quad-A^{\pi N(+)}(\nu=0,\nu_B=-1/2M,-1). \end{aligned}$$

In other words, we add and subtract the quantity $A^{\pi N(+)}(0,-1/2M,-1)$. In the difference term D, we evaluate $A^{\pi N(+)}(0,-1/2M,-1)$ by using the fixed momentum transfer dispersion relation for $A^{\pi N(+)}$. [See Eq. (38)] with a broad area subtraction. This is just the method used above to evaluate $A^{\pi N(+)}(0,0,-1)$. The results are shown in Table I. Clearly, in forming the difference D of the amplitudes for different values of momentum transfer ν_B, much of the variation of the result with ω_m cancels out. This is probably not accidental. If we take as error estimate the variation of D as ω_m is moved across the (3,3) resonance peak, we find

$$D=2.65\pm0.3. \qquad (50)$$

We now add back $A^{\pi N(+)}(0,-1/2M,-1)$ evaluated by an independent method. Let us recall that $\nu_B=-1/(2M)$ corresponds to forward pion-nucleon scattering. Since the even isotopic spin forward scattering amplitude is given by

$$F^{(+)}(\nu)=A^{\pi N(+)}(\nu,-1/2M,-1)+\nu B^{\pi N(+)}(\nu,-1/2M,-1), \qquad (51)$$

we have

$$F^{(+)}(0)=A^{\pi N(+)}(0,-1/2M,-1). \qquad (52)$$

Thus, we can use ordinary forward dispersion relations[13] to evaluate $A^{\pi N(+)}(0,-1/2M,-1)$. Making a broad area subtraction gives

$$\begin{aligned} A^{\pi N(+)}&(0,-1/2M,-1)\\ &=\frac{g_r^2}{M}\frac{\nu_m}{[(\nu_m^2-1/4M^2)(1-1/4M^2)]^{1/2}}\\ &+\frac{2}{\pi}\int_1^{\nu_m}\frac{d\nu'}{\nu'}\frac{\mathrm{Re}F^{(+)}(\nu')\nu_m}{[(\nu'-1)(\nu'+1)(\nu_m-\nu')(\nu_m+\nu')]^{1/2}}\\ &-\frac{2}{\pi}\int_{\nu_m}^{\infty}\frac{d\nu'}{\nu'}\frac{\mathrm{Im}F^{(+)}(\nu')\nu_m}{[(\nu'-1)(\nu'+1)(\nu'-\nu_m)(\nu'+\nu_m)]^{1/2}}. \end{aligned} \qquad (53)$$

We recall that

$$\mathrm{Re}F^{(+)}(\nu')=\frac{4\pi}{M}[2M\nu'+M^2+1]^{1/2}\,\mathrm{Re}(f_1^{(+)}+f_2^{(+)})', \qquad (54)$$

where $f_1^{(+)}$ and $f_2^{(+)}$ are the usual center-of-mass [isospin (+)] pion-nucleon scattering amplitudes. Furthermore,[13]

$$\mathrm{Im}F^{(+)}(\nu')=\tfrac{1}{2}(\nu'^2-1)^{1/2}[\sigma_+(\nu')+\sigma_-(\nu')], \qquad (55)$$

where $\sigma_+(\nu')$ and $\sigma_-(\nu')$ are, respectively, the total π^+p and π^-p cross sections. To evaluate the integrals, we used Roper's phase shifts for laboratory pion kinetic energies below 700 MeV. Above 700 MeV, we used the tabulation of σ_+ and σ_- given by Amblard *et al.*[14] and the

[13] For example, see the article by J. D. Jackson in *Dispersion Relations*, edited by G. R. Screaton (Interscience Publishers, Inc., New York, 1961), p. 38.
[14] B. Amblard *et al.*, Phys. Letters 10, 138 (1964).

TABLE II. $A^{\pi N(+)}$ versus length of the square root cut, as calculated by using forward scattering dispersion relations.

ω_m	1.5	2.1	2.7	3.3	3.9
$A^{\pi N(+)}(0,-1/2M,-1)$	26.33	26.23	26.15	26.09	26.07

asymptotic region fit of Von Dardel *et al.*[15] The results, shown in Table II, give

$$A^{\pi N(+)}(0,-1/2M,-1)=26.15\pm0.2,\quad A^{\pi N(+)}(0,0,-1)=28.8\pm0.4, \tag{56}$$

where we have taken as the error estimate the variation of $A^{\pi N(+)}(0,-1/2M,-1)$ as ω_m is varied from 1.5 to 3.9. We have not included in the error estimate the error in the factor g_r^2/M appearing in Eq. (53), since when we divide by g_r^2/M to compare the left- and right-hand sides of Eq. (31) this error drops out.

The values of $A^{\pi N(+)}(0,-1/2M,-1)$ obtained by using fixed momentum transfer dispersion relations (Table I) and forward scattering dispersion relations (Table II) are in excellent agreement. When fixed momentum transfer dispersion relations are used, the total result for $A^{\pi N(+)}(0,-1/2M,-1)$ comes from the integration over the physical cut. By contrast, when forward scattering dispersion relations are used, nearly all of the total comes from the pole term in the dispersion relations, which leads to the term $(g_r^2/M)\nu_m[(\nu_m^2-1/4M^2)\times(1-1/4M^2)]^{-1/2}$ in Eq. (53). Thus, the two methods "sample" pion-nucleon scattering in very different ways. Their agreement gives us confidence that the numbers obtained from the dispersion relations calculations are reliable.

B. Model for Going Off Mass Shell in k^2

In order to compare the consistency condition with experiment we must calculate the difference

$$[A^{\pi N(+)}(0,0,0)/K^{NN\pi}(0)]-A^{\pi N(+)}(0,0,-1). \tag{57}$$

To motivate the model which we use, let us return for a moment to the fixed momentum transfer dispersion relation for $A^{\pi N(+)}(\nu,0,-1)$,

$$A^{\pi N(+)}(\nu,0,-1)=\frac{1}{\pi}\int_{\nu_0}^{\infty}d\nu'\,\mathrm{Im}A^{\pi N(+)}(\nu',0,-1)\times\left[\frac{1}{\nu'-\nu}+\frac{1}{\nu'+\nu}\right]. \tag{58}$$

Let us proceed as if no subtractions were necessary. We evaluate the integral by keeping only the resonant (3,3) state in the integrand and going to the static limit. This gives

$$A^{\pi N(+)}(0,0,-1)\approx\frac{32}{3}M\pi\cdot\frac{1}{\pi}\int_1^{\infty}d\omega\frac{\mathrm{Im}f_{3,3}}{|\mathbf{q}|^2}, \tag{59}$$

where $|\mathbf{q}|$ is the pion center-of-mass momentum and where $f_{3,3}$ is the resonant (3,3) partial wave amplitude. According to Chew *et al.*,[10] in the narrow resonance approximation one finds that

$$\frac{1}{\pi}\int_1^{\infty}d\omega\frac{\mathrm{Im}f_{3,3}}{|\mathbf{q}|^2}\approx\frac{g_r^2}{12\pi M^2}, \tag{60}$$

giving

$$A^{\pi N(+)}(0,0,-1)\approx(8/9)(g_r^2/M)\approx24.4. \tag{61}$$

This number is in good agreement with those obtained above by the proper procedure of using subtracted dispersion relations. The fact that a (3,3) dominant, unsubtracted dispersion relation calculation gives a reasonable result for $A^{\pi N(+)}(0,0,-1)$ suggests that such a calculation may also give a reasonable estimate of the change in $A^{\pi N(+)}$ produced by going off mass shell. Thus, as our model for going off mass shell in k^2, we take

$$\Delta\equiv[A^{\pi N(+)}(0,0,0)/K^{NN\pi}(0)]-A^{\pi N(+)}(0,0,-1),$$
$$=\frac{2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu'}{\nu'}\mathrm{Im}\left[\frac{A_{3,3}{}^{\pi N(+)}(\nu',0,0)}{K^{NN\pi}(0)}-A_{3,3}{}^{\pi N(+)}(\nu',0,-1)\right], \tag{62}$$

where the subscript 3, 3 indicates that only the resonant partial wave is to be retained.[16]

The integral in Eq. (62) can be evaluated once the off-mass-shell partial wave amplitude $f_{3,3}(\nu',k^2=0)$ is known. It turns out that in the (3,3) resonance region, a very good estimate of $f_{3,3}(\nu',k^2=0)$ is given by

$$f_{3,3}(\nu',k^2=0)\approx f_{3,3}(\nu',k^2=-1)\frac{f_{3,3}{}^B(\nu',k^2=0)}{f_{3,3}{}^B(\nu',k^2=-1)}, \tag{63}$$

where $f_{3,3}{}^B$ denotes the (3,3) projection of the Born approximation.[17] Roughly speaking, the reasons for the validity of Eq. (63) are:

(i) Equation (63) gives $f_{3,3}(\nu',k^2=0)$ the phase of the (3,3) on-mass-shell amplitude, as is required by unitarity.

(ii) The left hand, or "potential" singularity of $f_{3,3}(\nu',k^2)$ nearest to the physical cut is determined entirely by $f_{3,3}{}^B(\nu',k^2)$. Multiplying $f_{3,3}(\nu',-1)$ by $f_{3,3}{}^B(\nu',0)/f_{3,3}{}^B(\nu',-1)$ gives the right-hand side of Eq. (63) approximately the correct nearly potential singularity structure for $f_{3,3}(\nu',0)$. A detailed numerical analysis[18] indicates that the error involved in using Eq.

[15] G. von Dardel *et al.*, Phys. Rev. Letters **8**, 173 (1962).

[16] A justification for this model would be provided if one could prove that $\Delta(\nu)\equiv A^{\pi N(+)}(\nu,0,0)/K^{NN\pi}(0)-A^{\pi N(+)}(\nu,0,-1)$ satisfies an unsubtracted dispersion relation in the variable ν. Then $\Delta(0)$ could be expressed as an integral of $\mathrm{Im}\Delta(\nu)$ over the physical cut. Since only the (3,3) phase shift is appreciable at low energy, it would be reasonable to keep only the (3,3) partial wave in $\mathrm{Im}\Delta(\nu)$.

[17] E. Ferrari and F. Selleri, Nuovo Cimento **21**, 1028 (1961).

[18] S. L. Adler (to be published).

(63) for $f_{3,3}(\nu',0)$, in the (3,3) resonance region, may be as small as half a percent.

Since $f_{3,3}{}^{B}(\nu',0)$ is proportional to $K^{NN\pi}(0)$, the pionic form factor of the nucleon drops out of the calculation. Substituting Eq. (63) into Eq. (62) and doing the integration numerically gives the result

$$\Delta \approx -0.5. \tag{64}$$

Hence the model we have used indicates that extrapolation off mass shell has only a small effect, of order 2% of g_r^2/M. This figure corresponds to the fact that the two terms in the integrand of Eq. (62) cancel up to small terms of order M_π^2/M^2, which is about 2%. The need to use a model is unfortunate, and the extrapolation off mass shell is the least certain aspect of the comparison of Eq. (31) with experiment. However, the apparent smallness of Δ indicates that the model would have to fail very badly for there to be an appreciable effect on the numerical results.

C. Summary

Adding the −0.5 obtained from going off mass shell to the results of Subsection A gives the final results shown in Table III. They indicate that unless the model

TABLE III. Final results for $A^{\pi N(+)}(0,0,0)M/K^{NN\pi}(0)g_r^2$. The error estimates are obtained as indicated in the text.

Method	Result	Error estimate
Threshold subtraction, using Woolcock's S- and P-wave scattering lengths.	1.07	±0.04
Threshold subtraction, using Roper's phase shift fits for all scattering lengths.	1.20	···
Broad area subtraction, using fixed momentum transfer dispersion relations.	1.03	±0.04
Alternative broad area subtraction method, using forward scattering dispersion relations.	1.03	±0.015

used for going off mass shell is badly in error, the consistency condition of Eq. (31) is satisfied to within 10%, and quite possibly to within 5%. This fact, together with the success of the Goldberger-Treiman relation, suggests that the PCAC hypothesis deserves further study.

IV. OTHER CONSISTENCY CONDITIONS

The consistency condition on pion-nucleon scattering is not the only condition on the strong interactions which is implied by PCAC. In this section, we discuss briefly the conditions connected with several other scattering amplitudes.

A. Condition on Pion-Pion Scattering

Let us consider the pion-pion scattering reaction illustrated in Fig. 2.[19] Let (p_1,α), (p_2,β) be the four-momenta and isospin indices of the initial pions, and (p_3,α'), (p_4,β') the four-momenta and isospin indices of the final pions. We take all four-momenta to be incoming, so that the condition of energy-momentum conservation reads

$$p_1+p_2+p_3+p_4=0. \tag{65}$$

We introduce the standard Mandelstam variables s, t, u by

$$\begin{aligned} s&=(p_1+p_2)^2=(p_3+p_4)^2,\\ t&=(p_1+p_4)^2=(p_2+p_3)^2,\\ u&=(p_1+p_3)^2=(p_2+p_4)^2,\\ s+t+u&=p_1^2+p_2^2+p_3^2+p_4^2. \end{aligned} \tag{66}$$

The isospin structure of the pion-pion scattering matrix element is

$$(16p_{10}p_{20}p_{30}p_{40})^{1/2}\langle\pi\pi^{\text{out}}|\pi\pi^{\text{in}}\rangle = \psi_{\alpha'}{}^*\psi_{\beta'}{}^*M^{\pi\pi}(s,t,u)_{\alpha\beta,\alpha'\beta'}\psi_\alpha\psi_\beta. \tag{67}$$

From the requirement that the scattering amplitude be symmetric under interchange of the pions, we find that

$$M^{\pi\pi}(s,t,u)_{\alpha\beta,\alpha'\beta'}=A^{\pi\pi}(s|t,u)\delta_{\alpha\beta}\delta_{\alpha'\beta'}+A^{\pi\pi}(t|u,s)\delta_{\alpha\alpha'}\delta_{\beta\beta'} +A^{\pi\pi}(u|t,s)\delta_{\alpha\beta'}\delta_{\beta\alpha'}, \tag{68}$$

where $A^{\pi\pi}(s|t,u)$ is a symmetric function of t and u. $A^{\pi\pi}$ also depends on p_1^2, p_2^2, p_3^2, and p_4^2. It is easy to see that at the symmetric point $s=t=u=(p_1^2+p_2^2+p_3^2+p_4^2)/3$, $A^{\pi\pi}$ is left invariant by the interchange $p_1^2\leftrightarrow p_2^2$, by the interchange $p_3^2\leftrightarrow p_4^2$, and by the simultaneous interchanges $p_1^2\leftrightarrow p_3^2$, $p_2^2\leftrightarrow p_4^2$.

Let us now consider the axial-vector matrix element $\langle\pi\pi|J_\lambda{}^A|\pi\rangle$. Let $p_2\equiv k$ be the momentum transfer and β the isospin index associated with the current $J_\lambda{}^A$, while we take (p_1,α), (p_3,α'), and (p_4,β') to be the four-momenta and isospin indices of the three pions. The isospin structure of the axial-vector matrix element is

$$(8p_{10}p_{30}p_{40})^{1/2}\langle\pi\pi|J_\lambda{}^A|\pi\rangle = \psi_{\alpha'}{}^*\psi_{\beta'}{}^*M(s,t,u)_{\alpha\beta,\alpha'\beta'}{}^\lambda\psi_\alpha\psi_\beta{}^+. \tag{69}$$

Defining Mandelstam variables as above, we find that the amplitude $M(s,t,u)_{\alpha\beta,\alpha'\beta'}{}^\lambda$ is given by

$$\begin{aligned} M(s,t,u)_{\alpha\beta,\alpha'\beta'}{}^\lambda &=[A_1(s|t,u)(p_3+p_4)_\lambda+A_2(s|t,u)(p_3-p_4)_\lambda\\ &\quad+A_3(s|t,u)p_{1\lambda}]\delta_{\alpha\beta}\delta_{\alpha'\beta'}+[A_1(t|u,s)(p_1+p_3)_\lambda\\ &\quad+A_2(t|u,s)(p_1-p_3)_\lambda+A_3(t|u,s)p_{4\lambda}]\delta_{\alpha\alpha'}\delta_{\beta\beta'}\\ &\quad+[A_1(u|t,s)(p_1+p_4)_\lambda+A_2(u|t,s)(p_1-p_4)_\lambda\\ &\quad+A_3(u|t,s)p_{3\lambda}]\delta_{\alpha\beta'}\delta_{\beta\alpha'}, \end{aligned} \tag{70}$$

where $A_1(s|t,u)$ and $A_3(s|t,u)$ are symmetric functions and $A_2(s|t,u)$ is an antisymmetric function of the variables t and u.

There are no pole terms which contribute to the amplitudes A_1, A_2, and A_3 of Eq. (70). Thus, when

$$p_2\cdot p_1=p_2\cdot p_3=p_2\cdot p_4=0, \tag{71}$$

[19] G. F. Chew and S. Mandelstam, Phys. Rev. 119, 467 (1960).

we have $p_{2\lambda}M(s,t,u)_{\alpha\beta,\alpha'\beta'}{}^{\lambda}=0$, in other words,

$$\langle\pi\pi|\partial_\lambda J_\lambda{}^A|\pi\rangle=0. \qquad (72)$$

Equation (71) implies that we are at the symmetric point

$$s=t=u=k^2-M_\pi{}^2. \qquad (73)$$

Since $s+t+u=-3M_\pi{}^2+k^2$, we see that Eq. (73) can be satisfied when $k^2=0$, giving the result that $\langle\pi\pi|\partial_\lambda J_\lambda{}^A|\pi\rangle$ vanishes when $k^2=0$ and $s=t=u=-M_\pi{}^2$. The PCAC hypothesis allows us to write

$$\langle\pi\pi|\partial_\lambda J_\lambda{}^A|\pi\rangle=C\langle\pi\pi|\varphi_\pi|\pi\rangle. \qquad (74)$$

Consequently, PCAC implies that

$$A^{\pi\pi}(s=-M_\pi{}^2|t=-M_\pi{}^2,u=-M_\pi{}^2|k^2=0)=0, \qquad (75)$$

where $-k^2$ is the (mass)^2 of one of the four pions and where the other three pions are on mass shell.

Comparison of Eq. (75) with experiment will be difficult, since the effect of one of the pions being off mass shell is very likely to be important. In particular, the negative of the pion-pion amplitude at the on-mass-shell symmetric point,

$$-A^{\pi\pi}(s=-\tfrac{4}{3}M_\pi{}^2|t=-\tfrac{4}{3}M_\pi{}^2,u=-\tfrac{4}{3}M_\pi{}^2|k^2=-M_\pi{}^2) \qquad (76)$$

is just the effective pion-pion coupling constant[19] and is not zero.

B. Condition on Pion-Lambda Scattering[20]

The derivation in this case closely parallels the derivation given in Sec. II for the condition on πN scattering. The generalized Born approximation diagrams for $\langle\pi\Lambda|J_\lambda{}^A|\Lambda\rangle$ are shown in Fig. 3. In the derivation of Sec. II, we make the replacements

$$ig_r\bar\psi_N\gamma_5\boldsymbol{\tau}\psi_N\cdot\boldsymbol{\varphi}_\pi\rightarrow ig_{\Lambda\Sigma}(\bar\psi_{\Sigma^-}\gamma_5\psi_\Lambda+\bar\psi_\Lambda\gamma_5\psi_{\Sigma^+})\varphi_{\pi^+}+\cdots \qquad (77)$$

to define the $\Lambda\Sigma\pi$ strong vertex[21];

$$g_A\bar\psi_N\gamma_\lambda\gamma_5\tau^+\psi_N\rightarrow g_A{}^{\Lambda\Sigma}(\bar\psi_{\Sigma^+}\gamma_\lambda\gamma_5\psi_\Lambda+\bar\psi_\Lambda\gamma_\lambda\gamma_5\psi_{\Sigma^-})+\cdots \qquad (78)$$

to define the $\Lambda\Sigma$ weak vertex; and

$$A_{\alpha\beta}{}^{\pi N}-ikB_{\alpha\beta}{}^{\pi N}\rightarrow(A^{\pi\Lambda}-ikB^{\pi\Lambda})\delta_{\alpha\beta} \qquad (79)$$

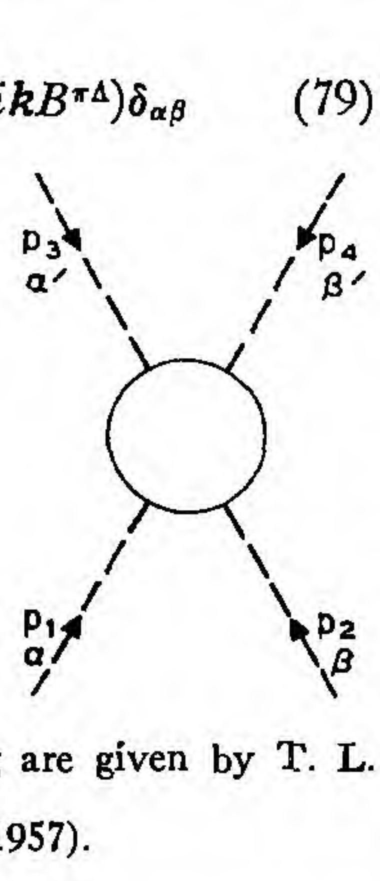


FIG. 2. Four-momenta and isospin indices for $\pi\pi$ scattering.

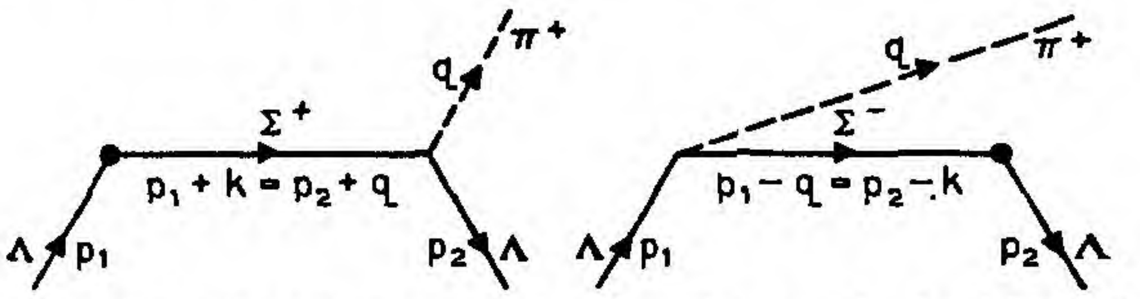


FIG. 3. Generalized Born approximation diagrams for $\langle\pi\Lambda|J_\lambda{}^A|\Lambda\rangle$.

to define the $\pi\Lambda$ scattering amplitudes. Equation (27) becomes

$$-2M\nu(\bar A_1+\bar A_2)+2M\nu_B\bar A_3+2g_{\Lambda\Sigma}g_A{}^{\Lambda\Sigma}(0)\times\left\{1-\frac{\sigma}{2}\left[\frac{1}{\nu_B-\nu+\sigma}+\frac{1}{\nu_B+\nu+\sigma}\right]\right\}=\frac{g_A{}^{\Lambda\Sigma}(0)(M_\Lambda+M_\Sigma)}{g_{\Lambda\Sigma}K^{\Lambda\Sigma}(0)}A^{\pi\Lambda}(\nu,\nu_B,k^2=0), \qquad (80)$$

where $\sigma=(M_\Sigma{}^2-M_\Lambda{}^2)/(2M_\Lambda)$, $\nu=-(p_1+p_2)\cdot k/(2M_\Lambda)$, $\nu_B=q\cdot k/(2M_\Lambda)$, and where $K^{\Lambda\Sigma}$ is the form factor of the $\Lambda\Sigma\pi$ vertex, normalized so that $K^{\Lambda\Sigma}(-M_\pi{}^2)=1$. Setting $\nu=\nu_B=0$ gives the consistency condition

$$0=A^{\pi\Lambda}(\nu=0,\nu_B=0,k^2=0). \qquad (81)$$

This is a null condition and thus differs greatly from the condition derived for πN scattering. The difference arises from the fact that the intermediate state baryon in the generalized Born approximation for $\langle\pi\Lambda|J_\lambda{}^A|\Lambda\rangle$ is a Σ, which has a mass unequal to that of the external Λ. This makes the quantity σ in Eq. (80) different from zero, with the result that the coefficient of $2g_{\Lambda\Sigma}g_A{}^{\Lambda\Sigma}(0)$ vanishes when ν and ν_B are set equal to zero. In the case of πN scattering, σ is zero, and a nonnull condition on $A^{\pi N}$ is obtained. It would be an interesting problem to try to determine from a study of $\pi\Lambda$ scattering whether Eq. (81) is satisfied.

C. Other Reactions

The space-spin structures of $\langle\pi K|J_\lambda{}^A|K\rangle$ and $\langle\pi(\Sigma,\Xi)|J_\lambda{}^A|(\Sigma,\Xi)\rangle$ are similar to the space-spin structures of $\langle\pi\pi|J_\lambda{}^A|\pi\rangle$ and $\langle\pi N|J_\lambda{}^A|N\rangle$, respectively. Consequently, there will be consistency conditions on the πK, the $\pi\Sigma$, and the $\pi\Xi$ scattering amplitudes. Since $\langle\pi(\Sigma,\Xi)|J_\lambda{}^A|(\Sigma,\Xi)\rangle$ has a generalized Born approximation diagram with an intermediate (Σ,Ξ), the consistency condition will be a nonnull condition, like Eq. (31) for πN scattering, rather than a null condition, like Eq. (81) for $\pi\Lambda$ scattering.

We have not studied reactions with more than two particles in the final state. It would be interesting, for example, to determine from a study of $\langle\pi\pi N|J_\lambda{}^A|N\rangle$ whether PCAC implies a consistency condition involving the amplitudes for $\pi+N\rightarrow\pi+\pi+N$.

ACKNOWLEDGMENTS

I wish to thank Professor S. B. Treiman for discussions which greatly helped to clarify the ideas presented

[20] References dealing with $\pi\Lambda$ scattering are given by T. L. Trueman, Phys. Rev. **127**, 2240 (1962).
[21] M. Gell-Mann, Phys. Rev. **106**, 1296 (1957).

in this paper. I also wish to thank Dr. D. G. Cassel and Dr. C. G. Callan for help with the computer work, Dr. P. Kantor for an interesting discussion, and Dr. J. R. Klauder for reading the manuscript. I am grateful to Dr. L. D. Roper for supplying the pion-nucleon phase shifts used in the numerical work, and to the National Science Foundation for a predoctoral fellowship during the years 1961–1964.

APPENDIX

We derive here the equations used in the numerical calculations described in Sec. III. Let us consider the reaction $\pi(k)+N(p_1)\to\pi(q)+N(p_2)$, where the four-momenta of the particles are indicated in parentheses. We take the nucleons and the final pion to be on mass shell,

$$p_1^2=p_2^2=-M^2,\quad q^2=-M_\pi^2, \tag{A1}$$

but keep k^2 arbitrary. Let $\mathbf{k}=-\mathbf{p}_1$ and $\mathbf{q}=-\mathbf{p}_2$ be, respectively, the momenta of the initial and final pion in the center-of-mass frame of the reaction, and let k_0, p_{10}, q_0, p_{20} be the center-of-mass particle energies. We denote by W the total center-of-mass energy

$$\begin{aligned} W&=k_0+p_{10}=q_0+p_{20},\\ k_0&=(W^2-M^2-k^2)/2W,\\ q_0&=(W^2-M^2+M_\pi^2)/2W,\\ p_{10}&=(W^2+M^2+k^2)/2W,\\ p_{20}&=(W^2+M^2-M_\pi^2)/2W. \end{aligned} \tag{A2}$$

We denote by φ the center-of-mass scattering angle between the final and initial pion, so that

$$y\equiv\cos\varphi=\hat{q}\cdot\hat{k}, \tag{A3}$$

where $\hat{q}$ and $\hat{k}$ are unit vectors along the directions of the final and initial pion, respectively. The magnitudes $|\mathbf{q}|$ and $|\mathbf{k}|$ are clearly given by

$$|\mathbf{q}|=(q_0^2-M_\pi^2)^{1/2},\quad |\mathbf{k}|=(k_0^2+k^2)^{1/2}. \tag{A4}$$

The quantities ν and ν_B are related to W and $\cos\varphi$ by

$$\begin{aligned} \nu-\nu_B&=(W^2-M^2)/2M,\\ \nu_B&=(1/2M)[|\mathbf{q}||\mathbf{k}|\cos\varphi-q_0k_0]. \end{aligned} \tag{A5}$$

The variable ω, frequently used in going to the static limit, is defined by

$$\omega=W-M. \tag{A6}$$

Let us introduce center-of-mass amplitudes f_1 and f_2 by writing

$$\bar{u}(p_2)(A^{\pi N}-ik B^{\pi N})u(p_1) =\frac{4\pi W}{M}\chi_{s,f}^\dagger[f_1+f_2\sigma\cdot\hat{q}\sigma\cdot\hat{k}]\chi_{s,i}, \tag{A7}$$

where $A^{\pi N}$ and $B^{\pi N}$ are the covariant amplitudes used in the text and where $\chi_{s,f}$ and $\chi_{s,i}$ are the nucleon spinors. (We suppress isospin structure.) The transformation relating the amplitudes f_1, f_2 to the amplitudes $A^{\pi N}$, $B^{\pi N}$ is

$$\begin{aligned} \frac{f_1}{[(p_{10}+M)(p_{20}+M)]^{1/2}}&=\frac{1}{2W}\frac{A^{\pi N}}{4\pi}+\frac{W-M}{2W}\frac{B^{\pi N}}{4\pi},\\ \frac{f_2}{[(p_{10}-M)(p_{20}-M)]^{1/2}}&=\frac{-1}{2W}\frac{A^{\pi N}}{4\pi}+\frac{W+M}{2W}\frac{B^{\pi N}}{4\pi}. \end{aligned} \tag{A8}$$

The partial wave expansion of f_1 and f_2 is given by

$$\begin{aligned} f_1&=\sum_{l=0}^{\infty} f_{l+}P_{l+1}'(y)-\sum_{l=2}^{\infty} f_{l-}P_{l-1}'(y),\\ f_2&=\sum_{l=1}^{\infty}(f_{l-}-f_{l+})P_l'(y),\\ f_{l+}&=\frac{1}{2}\int_{-1}^{1}dy[f_2P_{l+1}(y)+f_1P_l(y)],\\ f_{l-}&=\frac{1}{2}\int_{-1}^{1}dy[f_2P_{l-1}(y)+f_1P_l(y)], \end{aligned} \tag{A9}$$

where $f_{l\pm}$ is the amplitude for the partial wave with orbital angular momentum l and total angular momentum $J=l\pm\frac{1}{2}$. The symmetric isospin amplitude $f_{l\pm}^{(+)}$ is given in terms of the isotopic spin $\frac{3}{2}$ and $\frac{1}{2}$ amplitudes by

$$f_{l\pm}^{(+)}=\tfrac{2}{3}f_{l\pm}^{(3/2)}+\tfrac{1}{3}f_{l\pm}^{(1/2)}. \tag{A10}$$

Finally, we need the inverse of Eq. (A8) for the amplitude $A^{\pi N}$,

$$\frac{A^{\pi N}}{4\pi}=\frac{(W+M)f_1}{[(p_{10}+M)(p_{20}+M)]^{1/2}}-\frac{(W-M)f_2}{[(p_{10}-M)(p_{20}-M)]^{1/2}}. \tag{A11}$$

A. Equations for Threshold Subtraction and Static Limit

Let us first consider the case when $k^2=-M_\pi^2$ and derive the equations used in the threshold subtraction and the static limit treatments of the dispersion relations. Below the two-pion threshold,

$$f_{l\pm}^{(I)}=\exp[i\delta_{l\pm}^{(I)}]\sin\delta_{l\pm}^{(I)}/|\mathbf{q}|, \tag{A12}$$

where $\delta_{l\pm}^{(I)}$ is the phase shift. The scattering length $a_{l\pm}^{(I)}$ is defined by

$$a_{l\pm}^{(I)}=\lim_{|\mathbf{q}|\to 0}\frac{f_{l\pm}^{(I)}}{|\mathbf{q}|^{2l}}. \tag{A13}$$

Using the facts that

$$\cos\varphi = [(2M\nu_B + M_\pi^2)/|\mathbf{q}|^2] + 1\,, \qquad \text{(A14)}$$

and that the leading term of $P_l'(y)$ for large y is

$$P_l'(y) \sim [l(2l)!/2^l(l!)^2]y^{l-1}\,, \qquad \text{(A15)}$$

we find that, at threshold,

$$[f_1^{(+)}]_T = \sum_{l=0}^{\infty} [\tfrac{2}{3}a_{l+}^{(3/2)} + \tfrac{1}{3}a_{l+}^{(1/2)}] \times \frac{(l+1)[2(l+1)]!}{2^{l+1}[(l+1)!]^2}[2M\nu_B + M_\pi^2]^l\,,$$

$$\left[\frac{f_2^{(+)}}{|\mathbf{q}|^2}\right]_T = \sum_{l=1}^{\infty} \{\tfrac{2}{3}[a_{l-}^{(3/2)} - a_{l+}^{(3/2)}] + \tfrac{1}{3}[a_{l-}^{(1/2)} - a_{l+}^{(1/2)}]\} \times \frac{l(2l)!}{2^l(l!)^2}[2M\nu_B + M_\pi^2]^{l-1}\,, \qquad \text{(A16)}$$

$$\frac{[A^{\pi N(+)}]_T}{4\pi} = \left(1 + \frac{1}{2M}\right)[f_1^{(+)}]_T - 2M\left[\frac{f_2^{(+)}}{|\mathbf{q}|^2}\right]_T.$$

When $\nu_B = 0$, this is just the result stated in Eq. (41) of the text.

The static limit of $A^{\pi N(+)}$ is easily derived. According to Eqs. (A9–A11), when all partial wave amplitudes except $f_{3,3} \equiv f_{1+}^{(3/2)}$ are neglected, $A^{\pi N(+)}$ is given by

$$\frac{A^{\pi N(+)}}{4\pi} = \left[3\frac{W+M}{p_{20}+M}\frac{2M\nu_B + q_0^2}{|\mathbf{q}|^2} + \frac{(W-M)(p_{20}+M)}{|\mathbf{q}|^2}\right]\frac{2}{3}f_{3,3}. \qquad \text{(A17)}$$

In the static limit, when $\nu_B = 0$, this is

$$A^{\pi N(+)} \approx \frac{16}{3}\frac{M\pi\omega}{|\mathbf{q}|^2}f_{3,3}.$$

Since in the static limit (when $\nu_B = 0$) $\nu \approx \omega$ and $\nu_0 \approx 1$, we have

$$\frac{2}{\pi}\int_{\nu_0}^{\infty}\frac{d\nu'}{\nu'}\,\mathrm{Im}A^{\pi N(+)}(\nu', 0, -1) \approx \frac{32}{3}M\pi\cdot\frac{1}{\pi}\int_1^{\infty} d\omega\frac{\mathrm{Im}f_{3,3}}{|\mathbf{q}|^2}. \qquad \text{(A18)}$$

B. Equations for Extrapolation off Mass Shell

Now let us consider $k^2 \neq -M_\pi^2$ and derive the equations used for going off mass shell in k^2. According to our model, we wish to calculate

$$\mathrm{Im}\Delta(\nu,k^2) \equiv [\mathrm{Im}A_{3,3}^{\pi N(+)}(\nu,0,k^2)/K^{NN\pi}(k^2)] - \mathrm{Im}A_{3,3}^{\pi N(+)}(\nu, 0, -1) \qquad \text{(A19)}$$

at $k^2 = 0$. From Eqs. (A9–A11) and Eq. (63) of the text, $\mathrm{Im}\Delta(\nu,k^2)$ is given by

$$\mathrm{Im}\Delta(\nu,k^2) = \frac{4\pi}{|\mathbf{q}|^2}\frac{2}{3}\mathrm{Im}f_{3,3}\left[\frac{3(W+M)q_0^2}{p_{20}+M} + \omega(p_{20}+M)\right](L-1)\,, \qquad \text{(A20)}$$

with

$$L = \frac{1}{K^{NN\pi}(k^2)}\frac{f_{3,3}^B(\nu,k^2)}{f_{3,3}^B(\nu, k^2 = -M_\pi^2)}\frac{|\mathbf{q}|}{|\mathbf{k}|} \times \frac{\dfrac{3(W+M)q_0k_0}{[(p_{10}+M)(p_{20}+M)]^{1/2}} + \omega[(p_{10}+M)(p_{20}+M)]^{1/2}}{\dfrac{3(W+M)q_0^2}{p_{20}+M} + \omega(p_{20}+M)}. \qquad \text{(A21)}$$

The Born approximations are computed by substituting the isospin $\frac{3}{2}$ part of the Born approximation

$$B^{\pi NB(3/2)} = -g_r^2K^{NN\pi}(k^2)/|\mathbf{q}||\mathbf{k}|(y+a)\,,$$
$$a = (2p_{20}k_0 + k^2)/2|\mathbf{q}||\mathbf{k}|\,, \qquad \text{(A22)}$$

into Eq. (A8) to calculate $f_1^{B(3/2)}$ and $f_2^{B(3/2)}$. The $J = \frac{3}{2}$ projection is then done by using Eq. (A9). The result is

$$\frac{1}{K^{NN\pi}(k^2)}\frac{f_{3,3}^B(\nu,k^2)}{f_{3,3}^B(\nu, k^2 = -M_\pi^2)} = \frac{|\mathbf{q}|}{|\mathbf{k}|}\frac{N}{N'}\,, \qquad \text{(A23)}$$

$$N = \omega[(p_{10}+M)(p_{20}+M)]^{1/2}A(a) + (W+M)[(p_{10}-M)(p_{20}-M)]^{1/2}C(a)\,,$$

$$N' = \omega(p_{20}+M)A(a') + (W+M)(p_{20}-M)C(a')\,,$$

where

$$a' = (2p_{20}q_0 - M_\pi^2)/2|\mathbf{q}|^2\,,$$

$$A(a) = 1 - \frac{a}{2}\ln[(a+1)/(a-1)]\,, \qquad \text{(A24)}$$

$$C(a) = \frac{1}{2}\left[3a + \left(\frac{1-3a^2}{2}\right)\ln\left(\frac{a+1}{a-1}\right)\right].$$

Consistency Conditions on the Strong Interactions Implied by a Partially Conserved Axial-Vector Current, STEPHEN L. ADLER [Phys. Rev. **137**, B1022 (1965)]. In Eqs. (16) and (23),

$$[(p_{10}/M)(p_{20}/M)2k_0]^{1/2}$$

should be

$$[(p_{10}/M)(p_{20}/M)2q_0]^{1/2}.$$

PHYSICAL REVIEW VOLUME 139, NUMBER 6B 20 SEPTEMBER 1965

Consistency Conditions on the Strong Interactions Implied by a Partially Conserved Axial-Vector Current. II

STEPHEN L. ADLER*
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received 26 March 1965)

Consequences of the partially conserved axial-vector current (PCAC) hypothesis are explored. A set of simple rules is derived which relate the matrix element for any strong interaction process with the matrix element for the corresponding process in which an additional zero-mass, zero-energy pion is emitted or absorbed. A generalization to include lowest order electromagnetic processes is given. A theorem is stated and proved which shows how divergence equations of the form $\partial_\lambda J_\lambda = D$ are modified when a minimal electromagnetic interaction is switched on.

INTRODUCTION

In an earlier paper[1] it was shown that the hypothesis of partially conserved $\Delta S=0$ axial-vector current (PCAC) leads to consistency conditions involving solely the strong interactions. One of these conditions, relating the pion-nucleon scattering amplitude $A^{\pi N(+)}$ and the pion-nucleon coupling constant g_r, was shown to agree with experiment to within 10%. In this note we give a simplified and generalized derivation of the consistency conditions implied by PCAC. We will derive a set of simple rules which relate the matrix element for any strong interaction or first-order electromagnetic process with the matrix element for the corresponding process in which an additional zero-mass, zero-energy pion is emitted or absorbed. The rules are closely connected with the "chirality conservation" formulas of Nambu, Lurié, and Shrauner.

Let us begin by recalling certain definitions from (I). We denote by $J_\lambda{}^A$ the strangeness-conserving weak axial current. By partially conserved axial-vector current we mean the hypothesis that

$$\partial_\lambda J_\lambda{}^A = \frac{-i\sqrt{2}M_N M_\pi{}^2 g_A{}^N(0)}{g_r K^{NN\pi}(0)}\phi_\pi + R. \tag{1}$$

Here M_N is the nucleon mass, M_π is the pion mass, $g_A{}^N(0)$ is the β-decay axial-vector coupling constant $[g_A{}^N(0)\approx 1.2\times 10^{-5}/M_N{}^2]$, g_r is the rationalized, renormalized pion-nucleon coupling constant ($g_r{}^2/4\pi \approx 14.6$), and ϕ_π is the renormalized field operator which creates the π^+. The quantity $K^{NN\pi}(0)$ is the pionic form factor of the nucleon evaluated at zero virtual pion mass; $K^{NN\pi}$ is normalized so that $K^{NN\pi}(-M_\pi{}^2)=1$. In order to give content to the definition, we must specify properties of the residual operator R. We suppose that for states $\langle\beta(p_F)|$ and $|\alpha(p_I)\rangle$ for which $\langle\beta|\phi_\pi|\alpha\rangle\neq 0$, and for momentum transfer near the one pion pole at $-M_\pi{}^2$ [say, for $-M_\pi{}^2<(p_F-p_I)^2<M_\pi{}^2$], the matrix element of R is much smaller than the matrix element of the pion operator term. In other words, we postulate that if $\langle\beta|\phi_\pi|\alpha\rangle\neq 0$ and if $|(p_F-p_I)^2|<M_\pi{}^2$, then

$$\frac{|\langle\beta|R|\alpha\rangle|}{[\sqrt{2}M_N M_\pi{}^2 g_A{}^N(0)/g_r K^{NN\pi}(0)]|\langle\beta|\phi_\pi|\alpha\rangle|}\ll 1. \tag{2}$$

In what follows we derive equalities which hold rigorously if the residual operator R is zero. If R is not zero, but satisfies the inequality of Eq. (2), the "equals" signs should be replaced by "approximately equals" signs.

It will be helpful to introduce a number of abbreviations and definitions. We denote by k the momentum transfer p_F-p_I. Let us introduce the isotopic vector quantities $J_\lambda{}^{Aa}$, $\phi_\pi{}^a$ $(a=1, 2, 3)$, in terms of which

$$J_\lambda{}^A=\tfrac{1}{2}(J_\lambda{}^{A1}+iJ_\lambda{}^{A2}), \quad \phi_\pi=(1/\sqrt{2})(\phi_\pi{}^1+i\phi_\pi{}^2). \tag{3}$$

We denote the product $g_r\, K^{NN\pi}(0)$ by $g_r{}^{\pi N}(0)$. Then the generalization of Eq. (1) to all three isospin components $J_\lambda{}^{Aa}$ is (neglecting R)

$$\partial_\lambda J_\lambda{}^{Aa} = -i(2M_N M_\pi{}^2 g_A{}^N(0)/g_r{}^{\pi N}(0))\phi_\pi{}^a. \tag{4}$$

It will be convenient to introduce an isospin notation for the Σ and for the Ξ analogous to that for the nucleon N. We introduce isospinors and isospin column vectors as follows:

$$\Xi^0\to\begin{pmatrix}1\\0\end{pmatrix},\quad \Xi^-\to\begin{pmatrix}0\\1\end{pmatrix};$$

$$\Sigma^+\to\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\\0\end{bmatrix},\quad \Sigma^0\to\begin{bmatrix}0\\0\\1\end{bmatrix},\quad \Sigma^-\to\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\\0\end{bmatrix}. \tag{5}$$

By u_Ξ or u_Σ we will mean the ordinary Dirac spinor for the hyperon, multiplied by the appropriate isospinor or isospin column vector. Let τ^a denote the usual Pauli matrices, and let t^{Na}, $t^{\Xi a}$, and $t^{\Sigma a}$ be the matrices defined by

$$t^{Na}=t^{\Xi a}=\tau^a, \tag{6}$$

$$[t^{\Sigma a}]_{bc}=i\epsilon_{bca}. \tag{7}$$

Then we may write the baryon matrix elements of $J_\lambda{}^{Aa}$ and of $J_\pi{}^a=(-\Box+M_\pi{}^2)\phi_\pi{}^a$ as follows. (We omit the induced pseudoscalar terms in $J_\lambda{}^{Aa}$, since these are

* Junior Fellow, Society of Fellows.
[1] Stephen L. Adler, Phys. Rev. 137, B1022 (1965). We will refer to this paper as (I).

treated separately in the derivation below. See Refs. 4 and 6.)

$$
\begin{aligned}
\langle B(p_F)|J_\lambda{}^{Aa}|B(p_I)\rangle
&=\left(\frac{M_B}{p_{F0}}\frac{M_B}{p_{I0}}\right)^{1/2}\bar{u}_B(p_F)g_A{}^B\gamma_\lambda\gamma_5 t^{Ba}u_B(p_I)\,,\\
\langle B(p_F)|J_\pi{}^a|B(p_I)\rangle
&=\left(\frac{M_B}{p_{F0}}\frac{M_B}{p_{I0}}\right)^{1/2}\bar{u}_B(p_F)ig_r{}^{\pi B}\gamma_5 t^{Ba}u_B(p_I)\,.
\end{aligned}
\tag{8}
$$

Here B denotes N, Σ, or Ξ.

Using these definitions of the coupling constants, and Eq. (4), it is an easy matter to see that

$$
\frac{M_N g_A{}^N(0)}{g_r{}^{\pi N}(0)}=\frac{M_\Sigma g_A{}^\Sigma(0)}{g_r{}^{\pi\Sigma}(0)}=\frac{M_\Xi g_A{}^\Xi(0)}{g_r{}^{\pi\Xi}(0)}\,. \tag{9}
$$

Equation (9) will permit us to eliminate the axial-vector coupling constants $g_A{}^N$, $g_A{}^\Sigma$, and $g_A{}^\Xi$ from the consistency conditions obtained in the next section.

I. DERIVATION OF CONSISTENCY CONDITIONS

We take the matrix element of both sides of Eq. (4) between states $\langle\beta(p_F)^{\mathrm{out}}|$ and $|\alpha(p_I)^{\mathrm{in}}\rangle$, where β and α are any systems of strongly interacting particles. This gives

$$
\begin{aligned}
&k_\lambda\langle\beta(p_F)^{\mathrm{out}}|J_\lambda{}^{Aa}|\alpha(p_I)^{\mathrm{in}}\rangle\\
&=(2M_NM_\pi{}^2g_A{}^N(0)/g_r{}^{\pi N}(0))\langle\beta(p_F)^{\mathrm{out}}|\phi_\pi{}^a|\alpha(p_I)^{\mathrm{in}}\rangle,\\
&=\frac{2M_Ng_A{}^N(0)}{g_r{}^{\pi N}(0)}\frac{M_\pi{}^2}{M_\pi{}^2+k^2}\langle\beta(p_F)^{\mathrm{out}}|J_\pi{}^a|\alpha(p_I)^{\mathrm{in}}\rangle.
\end{aligned}
\tag{10}
$$

Let us examine what happens in the limit as $k\to 0$ $(p_F\to p_I)$. The right-hand side of Eq. (10) in most cases approaches a finite limit, since

$$
\lim_{p_F\to p_I}\langle\beta(p_F)^{\mathrm{out}}|J_\pi{}^a|\alpha(p_I)^{\mathrm{in}}\rangle \tag{11}
$$

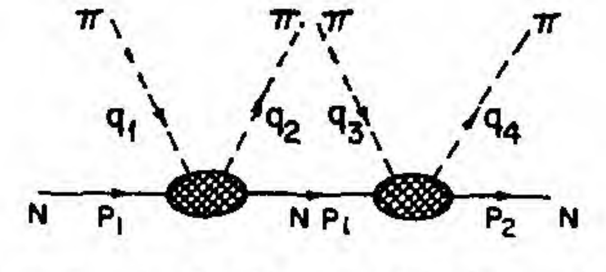


FIG. 1. The sort of situation which is excluded by the requirement that we avoid singularities of $\langle\beta^{\mathrm{out}}|\alpha^{\mathrm{in}}\rangle$. When $p_i{}^2=(p_1+q_1-q_2)^2=-M_N{}^2$, the diagram illustrated is infinite because the nucleon propagator joining the two bubbles is infinite. Such infinities can arise in general from pole diagrams contributing to $\langle\beta^{\mathrm{out}}|\alpha^{\mathrm{in}}\rangle$. (Pole diagrams are those which can be divided into two disconnected parts by cutting a single internal line.) We restrict ourselves in the text to values of the external four-momenta for which all pole diagrams contributing to $\langle\beta^{\mathrm{out}}|\alpha^{\mathrm{in}}\rangle$ are nonsingular.

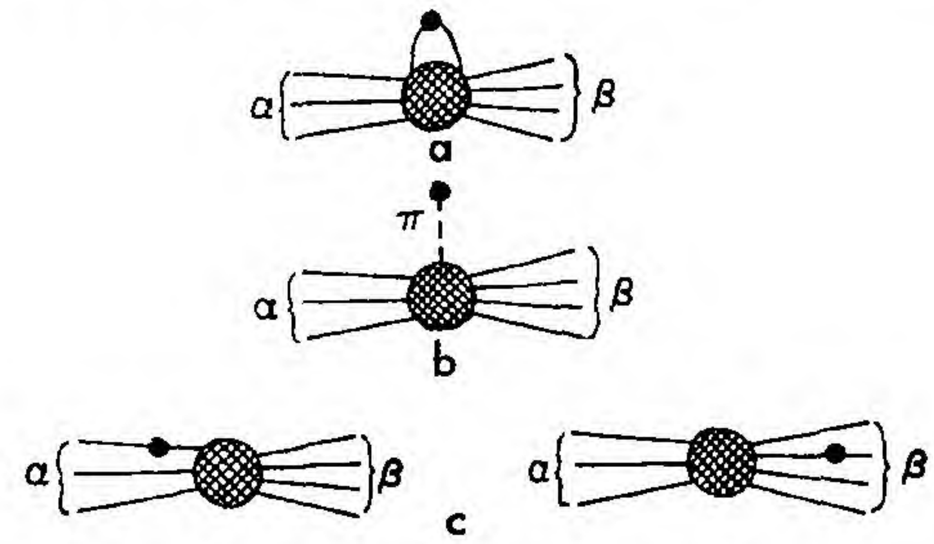


FIG. 2. Ways of attaching the proper vertex of $J_\lambda{}^A$, represented by a heavy dot. The proper vertex can be (a) attached to an internal line, (b) attached to a terminating external pion line, (c) attached to a nonterminating external line.

is just the matrix element for

$$
\alpha\to\beta+(\text{zero-mass, zero-energy pion}),
$$

and is in general nonzero.[2] Thus, the matrix element $\langle\beta(p_F)^{\mathrm{out}}|J_\lambda{}^{Aa}|\alpha(p_I)^{\mathrm{in}}\rangle$ must contain pole terms which go as $1/k$, in order that the scalar product of k with this matrix element have a finite limit. Clearly, if we can develop a simple set of rules for calculating these pole terms, we can calculate $\langle\beta(p_F)^{\mathrm{out}}|J_\pi{}^a|\alpha(p_I)^{\mathrm{in}}\rangle$ to zeroth order in k.

Calculation of the pole terms in $\langle\beta(p_F)^{\mathrm{out}}|J_\lambda{}^{Aa}|\alpha(p_I)^{\mathrm{in}}\rangle$ turns out to be quite easy. Let us restrict ourselves to values of the momenta of the particles in α and in β for which the matrix element $\langle\beta^{\mathrm{out}}|\alpha^{\mathrm{in}}\rangle$ has no singularities. (The sort of situation we wish to exclude is illustrated in Fig. 1.) The renormalized matrix element for $\langle\beta(p_F)^{\mathrm{out}}|J_\lambda{}^{Aa}|\alpha(p_I)^{\mathrm{in}}\rangle$ is obtained as follows[3]: First we write down a complete set of irreducible or "skeleton" diagrams for the matrix element. Then we make a series of insertions in the skeleton diagrams. We replace each bare propagator by the renormalized propagator, each bare strong-interaction vertex by the renormalized proper strong-interaction vertex, and each bare vertex where $J_\lambda{}^A$ acts by the renormalized proper vertex of $J_\lambda{}^A$.[4] We can divide the diagrams so obtained into three categories, according to where the proper vertex of $J_\lambda{}^A$ is attached: (a) The proper vertex of $J_\lambda{}^A$ is attached to an *internal* line [Fig. 2(a)]; (b) the proper vertex of $J_\lambda{}^A$ is attached to an *external pion* line which terminates [Fig. 2(b)]; (c) the proper vertex of $J_\lambda{}^A$ is attached to an *external line* which does not terminate [Fig. 2(c)].

[2] Note that the value of the limit depends in general on the *direction* in which k approaches zero.

[3] Let us review some definitions. The *skeleton* of a diagram is obtained by replacing all vertex parts by bare vertices and by omitting all self-energy parts from the propagators, so that only bare propagators appear. An *irreducible* or *"skeleton" diagram* is a diagram which is identical with its own skeleton. A *proper* vertex diagram is one which cannot be divided into two disconnected diagrams by cutting a single internal line.

[4] Note that the dominant part of the induced pseudoscalar coupling arises from the diagrams which give the one-pion pole term in dispersion theory. These diagrams are *improper* when considered as baryon-$J_\lambda{}^A$ vertices, and thus are not included in the proper baryon vertices of $J_\lambda{}^A$.

Corresponding to this division, we can write

$$\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda^{Aa}|\alpha(p_I)^{\rm in}\rangle$$
$$=\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm INT}$$
$$+\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm PION}$$
$$+\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm EXT}. \quad (12)$$

We now analyze in turn the contribution of each of the terms in Eq. (12):

(a) First let us consider the case where the proper vertex of $J_\lambda{}^A$ is attached to an internal line. Each diagram contributing to $\langle\beta(p_F)^{\rm out}|J_\lambda{}^A|\alpha(p_I)^{\rm in}\rangle^{\rm INT}$ corresponds to a diagram for $\langle\beta^{\rm out}|\alpha^{\rm in}\rangle$, but has an additional internal propagator. The requirement that $\langle\beta^{\rm out}|\alpha^{\rm in}\rangle$ be nonsingular means that all internal momenta are either integrated over or are off the mass shell. Thus the additional propagator cannot give rise to an infinity as $k\rightarrow 0$, and we conclude that $\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda{}^A|\alpha(p_I)^{\rm in}\rangle^{\rm INT}$ is of order k.[5]

(b) The sum of all diagrams where the proper vertex of $J_\lambda{}^A$ is attached to a terminating external pion line is proportional to

$$\langle\beta(p_F)^{\rm out}|J_\pi{}^c|\alpha(p_I)^{\rm in}\rangle[1/(k^2+M_\pi{}^2)]\langle\pi^c|J_\lambda{}^{Aa}|0\rangle. \quad (13)$$

Using Eq. (4) to evaluate $\langle\pi^c|J_\lambda{}^{Aa}|0\rangle$ gives the result

$$\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda{}^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm PION}$$
$$=\frac{-k^2}{k^2+M_\pi{}^2}\frac{2M_N g_A{}^N(0)}{g_r{}^{\pi N}(0)}\langle\beta(p_F)^{\rm out}|J_\pi{}^a|\alpha(p_I)^{\rm in}\rangle. \quad (14)$$

This is of order k^2 and may be neglected.[6]

(c) We next consider diagrams where the proper vertex of $J_\lambda{}^A$ is attached to a nonterminating external line. (We restrict ourselves to external lines of particles in the pseudoscalar meson or baryon octets.) These diagrams may be divided into two types, according to whether $J_\lambda{}^A$ changes or does not change the *mass* of the external particle.[7] The only case where the mass is changed is that where $J_\lambda{}^A$ changes an external Σ to a Λ or an external Λ to a Σ. Both of these cases make a contribution to $\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda{}^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm EXT}$ which is of order k, since the propagator which follows the proper vertex of $J_\lambda{}^A$ behaves as $(M_\Sigma{}^2-M_\Lambda{}^2)^{-1}$ as $k\rightarrow 0$, and thus is nonsingular. Finally, we will show that the diagrams where $J_\lambda{}^A$ is attached to a nonterminating external line, and does not change the mass, are of order k^{-1}. Insertion of $J_\lambda{}^A$ into a pseudoscalar meson line is forbidden by parity; insertion of $J_\lambda{}^A$ into a Λ line is forbidden by isospin. Thus, we need only consider insertions of $J_\lambda{}^A$ into external N, Σ, and Ξ lines. The contribution of the insertion of $J_\lambda{}^A$ into the line of a final baryon B of four-momentum p_B is

$$\left(\frac{M_B}{p_{B0}}\right)^{1/2}\bar{u}_B(p_B)g_A{}^B\gamma_\lambda\gamma_5 t^{Ba}\frac{1}{p_B-k-iM_B}\mathfrak{M}. \quad (15)$$

Here $\mathfrak{M}$ is the matrix element for the process $\alpha\rightarrow\beta$, with the final baryon B *virtual*. Since $p_B{}^2=-M_B{}^2$, the propagator can be written as

$$\frac{1}{p_B-k-iM_B}=\frac{p_B-k+iM_B}{-2p_B\cdot k+k^2}, \quad (16)$$

showing that there is indeed a singularity as $k\rightarrow 0$. To lowest order in k, we can neglect k in calculating $\mathfrak{M}$ and can retain only the term of order k^{-1} in Eq. (16). Thus, the insertion becomes

$$\left(\frac{M_B}{p_{B0}}\right)^{1/2}\bar{u}_B(p_B)g_A{}^B\gamma_\lambda\gamma_5 t^{Ba}\frac{p_B+iM_B}{-2p_B\cdot k}\mathfrak{M}(k=0). \quad (17)$$

Calculating $\mathfrak{M}$ with $k=0$ means that we keep the final baryon B on the mass shell. Furthermore, p_B+iM_B is just the positive frequency projection operator for B, with the property

$$(p_B+iM_B)p_B=(p_B+iM_B)iM_B. \quad (18)$$

Let us denote by $\mathfrak{M}^c$ the matrix element obtained by bringing all p_B in $\mathfrak{M}(k=0)$ to the left and replacing them by iM_B. Then the insertion becomes, finally,

$$\left(\frac{M_B}{p_{B0}}\right)^{1/2}\bar{u}_B(p_B)g_A{}^B\gamma_\lambda\gamma_5 t^{Ba}\frac{p_B+iM_B}{-2p_B\cdot k}\mathfrak{M}^c. \quad (19)$$

The crucial point is that

$$\langle\beta^{\rm out}|\alpha^{\rm in}\rangle=\delta_{\beta\alpha}+(2\pi)^4 i\delta(p_F-p_I)\mathfrak{M}(\alpha\rightarrow\beta), \quad (20a)$$

$$-i\mathfrak{M}(\alpha\rightarrow\beta)=\left(\frac{M_B}{p_{B0}}\right)^{1/2}\bar{u}_B(p_B)\mathfrak{M}^c \quad (20b)$$

[5] We assume, of course, that none of the proper vertices of $J_\lambda{}^A$ have a singularity as $k\rightarrow 0$.

[6] These diagrams form the dominant part of the induced pseudoscalar coupling. A statement much stronger than that they are of order k^2 can be made. Referring to Eq. (10), we note that the right-hand side may be written

$$(2M_N g_A{}^N(0)/g_r{}^{\pi N}(0))[1-k^2/(k^2+M_\pi{}^2)]\langle\beta(p_F)^{\rm out}|J_\pi{}^a|\alpha(p_I)^{\rm in}\rangle.$$

The part of this proportional to $k^2/(k^2+M_\pi{}^2)$ *exactly cancels* the contribution, given by Eq. (14), of the diagrams where $J_\lambda{}^A$ is attached to a terminating external pion line. Now $k^2/(k^2+M_\pi{}^2)$ has the property

$$\lim_{k\rightarrow 0}\lim_{M_\pi{}^2\rightarrow 0}k^2/(k^2+M_\pi{}^2)=1,$$
$$\lim_{M_\pi{}^2\rightarrow 0}\lim_{k\rightarrow 0}k^2/(k^2+M_\pi{}^2)=0,$$

whereas the terms in Eq. (12) labeled INT and EXT are independent of the order of the limiting operations:

$$\lim_{k\rightarrow 0}\lim_{M_\pi{}^2\rightarrow 0}\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda{}^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm INT,EXT}$$
$$=\lim_{M_\pi{}^2\rightarrow 0}\lim_{k\rightarrow 0}\langle\beta(p_F)^{\rm out}|k_\lambda J_\lambda{}^{Aa}|\alpha(p_I)^{\rm in}\rangle^{\rm INT,EXT}.$$

Hence the exact cancellation of terms proportional to $k^2/(k^2+M_\pi{}^2)$ means that the limit, as $M_\pi{}^2\rightarrow 0$, of the consistency conditions of Eq. (24) is identical with the consistency conditions which would be obtained in a theory in which the pion mass was set equal to zero at the outset. Note that by virtue of Eq. (4), in such a theory the axial-vector current would be exactly conserved.

[7] We are neglecting the electromagnetic interactions, so all particles in the same isospin multiplet are of equal mass.

is just the matrix element which describes the strong process $\alpha\rightarrow\beta$, with all particles on the mass shell. Thus, $\mathfrak{M}^c$ can be measured experimentally. Similar arguments show that the insertion of $J_\lambda{}^A$ into an initial baryon line gives

$$\left(\frac{M_B}{p_{B0}}\right)^{1/2}\mathfrak{M}'^c\frac{p_B+iM_B}{2p_B\cdot k}g_A{}^B\gamma_\lambda\gamma_5 t^{Ba}u_B(p_B),\qquad(21)$$

with

$$-i\mathfrak{M}(\alpha\rightarrow\beta)=(M_B/p_{B0})^{1/2}\mathfrak{M}'^c u_B(p_B).\qquad(22)$$

To sum up, we have analyzed the behavior of each of the terms in Eq. (12). Let us collect the results and write

$$\begin{aligned}(2M_N g_A{}^N(0)/g_r{}^{\pi N}(0))\langle\beta(p_F)^{\text{out}}|J_\pi{}^a|\alpha(p_I)^{\text{in}}\rangle+O(k^2)\\ =O(k)+O(k^2)\\ +\sum_{\substack{\text{external}\\ \text{lines}}}[\text{insertions in}-i\mathfrak{M}(\alpha\rightarrow\beta)]+O(k).\qquad(23)\end{aligned}$$

The three terms on the right-hand side of Eq. (23) refer, respectively, to the internal line, the terminating external pion line, and the nonterminating external line insertions of $J_\lambda{}^A$. Multiplying through by $g_r{}^{\pi N}(0)/[2M_N g_A{}^N(0)]$ and using Eq. (9) to eliminate the ratios $g_A{}^\Sigma(0)/g_A{}^N(0)$ and $g_A{}^\Xi(0)/g_A{}^N(0)$ in terms of strong-interaction coupling constants leads to the following set of rules:

$$\begin{aligned}\langle\beta(p_F)^{\text{out}}|J_\pi{}^a|\alpha(p_I)^{\text{in}}\rangle\\ =O(k)+\sum_{\substack{\text{external}\\ \text{lines}}}[\text{insertions in}-i\mathfrak{M}(\alpha\rightarrow\beta)].\qquad(24)\end{aligned}$$

Insertions

For external π, K, η, Λ, the insertion is zero. For external N, Σ, Ξ, denoted by B, the insertions are

final B:

$$\bar{u}_B(p_B)\rightarrow\bar{u}_B(p_B)\left[\frac{g_r{}^{\pi B}(0)}{2M_B}k\gamma_5 t^{Ba}\right]\frac{p_B+iM_B}{-2p_B\cdot k}\qquad(25\text{a})$$

initial B:

$$u_B(p_B)\rightarrow\frac{p_B+iM_B}{2p_B\cdot k}\left[\frac{g_r{}^{\pi B}(0)}{2M_B}k\gamma_5 t^{Ba}\right]u_B(p_B).\qquad(25\text{b})$$

These rules are the generalization to arbitrary processes of the consistency conditions derived in (I). It is an interesting fact that these rules are just what would be obtained if the effective pion-baryon coupling for pions with four-momentum near zero were pseudovector rather than pseudoscalar. This intimate connection between PCAC and gradient coupling theories was first noted by Feynman.[8]

As an illustration of the above rules, let us consider a special case. Let α be a single nucleon of four-momentum p_1 and any number of pions; similarly, let β be a single nucleon of four-momentum p_2 and any number of pions. Then we may write

$$\mathfrak{M}(\alpha\rightarrow\beta)=(M_N{}^2/p_{10}p_{20})^{1/2}\bar{u}_N(p_2)\mathfrak{N}u_N(p_1).\qquad(26)$$

According to the rules derived above,

$$\begin{aligned}&\langle\beta(p_F)^{\text{out}}|J_\pi{}^a|\alpha(p_I)^{\text{in}}\rangle\\ &=O(k)-\left(\frac{M_N{}^2}{p_{10}p_{20}}\right)^{1/2}i\bar{u}_N(p_2)\left\{\left[\frac{g_r{}^{\pi N}(0)}{2M_N}k\gamma_5\tau^a\right]\right.\\ &\quad\times\left[\frac{p_2+iM_N}{-2p_2\cdot k}\right]\mathfrak{N}+\mathfrak{N}\left[\frac{p_1+iM_N}{2p_1\cdot k}\right]\\ &\quad\left.\times\left[\frac{g_r{}^{\pi N}(0)}{2M_N}k\gamma_5\tau^a\right]\right\}u(p_1).\qquad(27)\end{aligned}$$

It is easily seen that Eq. (27) is equivalent to the "chirality conservation" formula obtained by Nambu and Lurié[9] in a theory in which the pion mass is zero and in which the axial-vector current is exactly conserved.[6] Nambu and Shrauner[10] and Shrauner[11] applied Eq. (27) to the case when $\alpha,\beta=\pi+N$ and found possible consistency with experiment. A simpler case was studied in (I), where we took $\alpha=N$, $\beta=\pi^b+N$. In this case $\mathfrak{N}$ is just the pion-nucleon vertex $ig_r\tau^b\gamma_5$ and $\langle(\pi^b N)^{\text{out}}|J_\pi{}^a|N^{\text{in}}\rangle$ is the pion-nucleon scattering amplitude. Introducing the usual pion-nucleon scattering-energy and momentum-transfer variables ν and ν_B,

$$\begin{aligned}p_1\cdot k&=-M_N(\nu-\nu_B),\\ p_2\cdot k&=-M_N(\nu+\nu_B),\end{aligned}\qquad(28)$$

we get from Eq. (27)

$$\begin{aligned}&\langle(\pi^b N)^{\text{out}}|J_\pi{}^a|N^{\text{in}}\rangle=\left(\frac{M_N{}^2}{p_{10}p_{20}}\right)^{1/2}K^{NN\pi}(0)\bar{u}_N(p_2)\\ &\quad\times\left\{\frac{g_r{}^2}{M_N}\delta_{ab}-ik\frac{g_r{}^2}{2M_N}\left[\frac{\tau^b\tau^a}{\nu_B-\nu}-\frac{\tau^a\tau^b}{\nu_B+\nu}\right]\right\}u_N(p_1).\qquad(29)\end{aligned}$$

The term $(g_r{}^2/M_N)\delta_{ab}$ leads to the consistency condition

$$\frac{A^{\pi N(+)}(\nu=0,\nu_B=0,k^2=0)}{K^{NN\pi}(0)}=\frac{g_r{}^2}{M_N},\qquad(30)$$

which was discussed in detail in (I).

II. MODIFICATION IN THE PRESENCE OF THE ELECTROMAGNETIC INTERACTIONS

It is interesting to see how the rules derived above are modified when the electromagnetic interactions are taken into account. Since isotopic spin is not a good

[8] R. P. Feynman, *Proceedings of the Aix-en-Provence International Conference on Elementary Particles* (Centre d'Etudes Nucléaires de Saclay, Seine et Oise, 1961), Vol. II, p. 210. I am very grateful to Dr. M. Veltman for calling my attention to this reference and for emphasizing the connection between PCAC and gradient coupling of the pion.

[9] Y. Nambu and D. Lurié, Phys. Rev. **125**, 1429 (1962); Y. Nambu and E. Shrauner, *ibid.* **128**, 862 (1962).

[10] Y. Nambu and E. Shrauner, Ref. 9.

[11] E. Shrauner, Phys. Rev. **131**, 1847 (1963).

quantum number in the presence of electromagnetism, we will work only with fields and currents with definite charge transformation properties. Thus, we replace the three equations contained in Eq. (4) by the equations

$$\partial_\lambda J_\lambda^{A(\pm)} = C\phi_\pi^{(\pm)},$$
$$\partial_\lambda J_\lambda^{A(0)} = \sqrt{2}C\phi_\pi^{(0)}, \tag{31}$$

where

$$J_\lambda^{A(\pm)} = \tfrac{1}{2}(J_\lambda^{A1} \mp iJ_\lambda^{A2}), \qquad J_\lambda^{A(0)} = J_\lambda^{A3};$$
$$\phi_\pi^{(\pm)} = (1/\sqrt{2})(\phi_\pi^1 \mp i\phi_\pi^2), \qquad \phi_\pi^{(0)} = \phi_\pi^3; \tag{32}$$
$$C = \frac{-i\sqrt{2}M_N M_\pi^2 g_A^N(0)}{g_r K^{NN\pi}(0)}.$$

[The superscript $(\pm)$ refers to the charge destroyed.] It is shown in the Appendix that to first order in the electric charge e $(e>0)$, the modification of Eqs. (31) in the presence of the electromagnetic interactions is

$$(\partial_\lambda \mp ieA_\lambda)J_\lambda^{A(\pm)} = C\phi_\pi^{(\pm)},$$
$$\partial_\lambda J_\lambda^{A(0)} = \sqrt{2}C\phi_\pi^{(0)}. \tag{33}$$

As is customary, A_λ denotes the electromagnetic field. Since all electromagnetic corrections to masses and coupling constants are of second order in e, questions such as whether to use the charged or neutral pion mass in computing C do not arise.

Equations (33) permit us to state a simple set of rules for computing (up to terms linear in the four-momentum of the added pion) the matrix elements $\langle\beta^{\text{out}}|J_\pi^{(\pm 0)}|(\alpha\gamma)^{\text{in}}\rangle$, where α and β are any systems of strongly interacting particles and where the initial photon γ may be real or virtual. The terms $\partial_\lambda J_\lambda^{A(\pm 0)}$ in Eqs. (33) give rise to insertions into the external baryon lines of $-i\mathfrak{M}(\alpha\gamma\rightarrow\beta)$ identical with those of Eq. (25), apart from trivial changes in the isospin factors arising from the use of fields and currents of definite charge. In addition, we must add to $\langle\beta^{\text{out}}|J_\pi^{(\pm)}|(\alpha\gamma)^{\text{in}}\rangle$ the term

$$\frac{\pm e g_r^{\pi N}(0)}{\sqrt{2}M_N g_A^N(0)}\langle\beta^{\text{out}}|A_\lambda J_\lambda^{A(\pm)}|(\alpha\gamma)^{\text{in}}\rangle \tag{34}$$

arising from the term $A_\lambda J_\lambda^{A(\pm)}$ in Eq. (33). Using the standard reduction formulas, we find to lowest order in e that

$$\langle\beta^{\text{out}}|A_\lambda(y)J_\lambda^{A(\pm)}(y)|(\alpha\gamma)^{\text{in}}\rangle$$
$$= \frac{\exp(ik'\cdot y)}{(2k_0')^{1/2}}\langle\beta^{\text{out}}|\epsilon_\lambda J_\lambda^{A(\pm)}(y)|\alpha^{\text{in}}\rangle, \tag{35}$$

where k' is the four-momentum and ϵ_λ the polarization four-vector of the photon γ. Equations (33), (34), and (35) allow us to calculate the matrix element for the emission of a zero-energy, zero-mass pion in photo- and electroproduction reactions. They are equivalent to the formalism derived for this purpose by Nambu and Shrauner,[9] who also discuss a detailed application to the reaction $e+N\rightarrow e+N+\pi$.

ACKNOWLEDGMENTS

I wish to thank Professor S. B. Treiman and Professor A. Pais for urging me to investigate further the question of PCAC consistency conditions. I have benefited from discussions with Dr. J. S. Bell, Professor M. A. B. Bég, Dr. C. Callan, Professor S. Coleman, Professor N. N. Khuri, Dr. M. Veltman, and Professor T. T. Wu.

APPENDIX

We give here a fairly general treatment of the way in which divergence equations of the form

$$\partial_\lambda J_\lambda = D \tag{A1}$$

are modified in the presence of electromagnetic interactions. We state the result in the form of a theorem.[12]

Theorem. Let ψ_j be the unrenormalized fields of particles of charge e_j. Let us consider a strong-interaction theory with the Lagrangian $\mathfrak{L}[\{\psi\},\{\partial_\sigma\psi\}]$, where $\{\psi\}$ denotes the set of the ψ_j. Let J_λ be a current with definite charge transformation properties (charge e_J) derived by making an infinitesimal gauge transformation on the fields ψ_j in the following manner[13]:

$$\psi_j \rightarrow \psi_j' = \psi_j + \Lambda F_j[\{\psi\}],$$
$$\mathfrak{L} \rightarrow \mathfrak{L}' = \mathfrak{L}[\{\psi'\},\{\partial_\sigma\psi'\}], \tag{A2}$$
$$J_\lambda = [\delta\mathfrak{L}'/\delta(\partial_\lambda\Lambda)]_{\Lambda=0}.$$

Then,

(1) In the absence of electromagnetic interactions the current J_λ satisfies

$$\partial_\lambda J_\lambda = D, \tag{A3}$$

with J_λ and D both functions of the ψ_j and the $\partial_\sigma\psi_j$ only:

$$J_\lambda = J_\lambda[\{\psi\},\{\partial_\sigma\psi\}],$$
$$D = D[\{\psi\},\{\partial_\sigma\psi\}]. \tag{A4}$$

(2) Inclusion of the electromagnetic interactions, with minimal electromagnetic coupling, changes Eqs. (A3) and (A4) to

$$(\partial_\lambda - ie_J A_\lambda)J_\lambda[\{\psi\},\{\pi_\sigma\}] = D[\{\psi\},\{\pi_\sigma\}], \tag{A5}$$

where $\pi_{j\sigma}$ denotes the quantity $(\partial_\sigma - ie_j A_\sigma)\psi_j$.

Proof. We proceed as if the fields were classical quantities, ignoring questions of commutation and anticommutation. Let us first consider the case when there are no electromagnetic interactions. The Lagrange equation of motion for the field ψ_j is

$$\frac{\delta\mathfrak{L}}{\delta\psi_j} = \partial_\sigma\frac{\delta\mathfrak{L}}{\delta(\partial_\sigma\psi_j)}. \tag{A6}$$

Under the gauge transformation

$$\psi_j \rightarrow \psi_j' = \psi_j + \Lambda F_j[\{\psi\}], \tag{A7}$$

[12] I am grateful to Professor S. Coleman for assistance in proving the theorem.

[13] M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960).

the derivatives $\partial_\sigma\psi_j$ and the Lagrangian $\mathcal{L}$ change according to

$$\begin{aligned}\partial_\sigma\psi_j\to\partial_\sigma\psi_j'&=\partial_\sigma\psi_j+(\partial_\sigma\Lambda)F_j+\Lambda(\partial_\sigma F_j),\\ \mathcal{L}\to\mathcal{L}'&=\mathcal{L}[\{\psi'\},\{\partial_\sigma\psi'\}]\\ &=\mathcal{L}[\{\psi+\Lambda F\},\{\partial_\sigma\psi+(\partial_\sigma\Lambda)F\\ &\qquad+\Lambda(\partial_\sigma F)\}].\end{aligned}\tag{A8}$$

From Eq. (A8) we find for the first variations,

$$\begin{aligned}\frac{\delta\mathcal{L}'}{\delta\Lambda}&=\sum_j\left[\frac{\delta\mathcal{L}'}{\delta\psi_j'}F_j+\frac{\delta\mathcal{L}'}{\delta(\partial_\sigma\psi_j')}\partial_\sigma F_j\right],\\ \frac{\delta\mathcal{L}'}{\delta(\partial_\lambda\Lambda)}&=\sum_j\frac{\delta\mathcal{L}'}{\delta(\partial_\lambda\psi_j')}F_j.\end{aligned}\tag{A9}$$

Eq. (A8) also implies that

$$\left[\frac{\delta\mathcal{L}'}{\delta\psi_j'}\right]_{\Lambda=0}=\frac{\delta\mathcal{L}}{\delta\psi_j},\quad\left[\frac{\delta\mathcal{L}'}{\delta(\partial_\lambda\psi_j')}\right]_{\Lambda=0}=\frac{\delta\mathcal{L}}{\delta(\partial_\lambda\psi_j)}.\tag{A10}$$

Together, Eqs. (A6), (A9), and (A10) imply that

$$\partial_\lambda\left[\frac{\delta\mathcal{L}'}{\delta(\partial_\lambda\Lambda)}\right]_{\Lambda=0}=\left[\frac{\delta\mathcal{L}'}{\delta\Lambda}\right]_{\Lambda=0}.\tag{A11}$$

We define

$$\begin{aligned}J_\lambda&\equiv[\delta\mathcal{L}'/\delta(\partial_\lambda\Lambda)]_{\Lambda=0},\\ D&\equiv[\delta\mathcal{L}'/\delta\Lambda]_{\Lambda=0};\end{aligned}\tag{A12}$$

these are clearly functions only of the $\{\psi\}$ and the $\{\partial_\sigma\psi\}$.

Let us now turn on the electromagnetic interactions. According to the hypothesis of minimal electromagnetic coupling, the Lagrangian is modified according to

$$\mathcal{L}\to\mathcal{L}^{\mathrm{EM}}=\mathcal{L}[\{\psi\},\{\pi_\sigma\}]+\mathcal{L}^{\mathrm{EM}0},\tag{A13}$$

where $\mathcal{L}^{\mathrm{EM}0}$ is the kinetic Lagrangian of the electromagnetic field A_σ and where $\pi_{j\sigma}$ is $(\partial_\sigma-ie_jA_\sigma)\psi_j$. The new Lagrange equation for the field ψ_j is

$$\partial_\sigma(\delta\mathcal{L}^{\mathrm{EM}}/\delta(\partial_\sigma\psi_j))=\delta\mathcal{L}^{\mathrm{EM}}/\delta\psi_j.\tag{A14}$$

Let us henceforth treat ψ_j and $\pi_{j\sigma}$, rather than ψ_j and $\partial_\sigma\psi_j$, as the independent variables in taking the variation of $\mathcal{L}^{\mathrm{EM}}$. Then the Lagrange equation becomes

$$\partial_\sigma\frac{\delta\mathcal{L}^{\mathrm{EM}}}{\delta\pi_{j\sigma}}=\frac{\delta\mathcal{L}^{\mathrm{EM}}}{\delta\psi_j}-ie_jA_\sigma\frac{\delta\mathcal{L}^{\mathrm{EM}}}{\delta\pi_{j\sigma}}.\tag{A15}$$

Now let us make the gauge transformation $\psi_j\to\psi_j'=\psi_j+\Lambda F_j$. The quantity $\pi_{j\sigma}$ and the Lagrangian $\mathcal{L}^{\mathrm{EM}}$ change according to

$$\begin{aligned}\pi_{j\sigma}\to\pi_{j\sigma}'&=\pi_{j\sigma}-ie_jA_\sigma\Lambda F_j+(\partial_\sigma\Lambda)F_j+\Lambda(\partial_\sigma F_j),\\ \mathcal{L}^{\mathrm{EM}}\to\mathcal{L}^{\mathrm{EM}\prime}&=\mathcal{L}[\{\psi'\},\{\pi_\sigma'\}]+\mathcal{L}^{\mathrm{EM}0}\\ &=\mathcal{L}[\{\psi+\Lambda F\},\{\pi_\sigma-ieA_\sigma\Lambda F\\ &\qquad+(\partial_\sigma\Lambda)F+\Lambda(\partial_\sigma F)\}]+\mathcal{L}^{\mathrm{EM}0}.\end{aligned}\tag{A16}$$

The first variations are

$$\begin{aligned}\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta\Lambda}&=\sum_j\left[\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta\psi_j'}F_j+\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta\pi_{j\sigma}'}(\partial_\sigma F_j-ie_jA_\sigma F_j)\right],\\ \frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta(\partial_\lambda\Lambda)}&=\sum_j\left[\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta\pi_{j\lambda}'}F_j\right].\end{aligned}\tag{A17}$$

Using the Lagrange equation, Eq. (A15), we see that

$$\partial_\lambda\left[\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta(\partial_\lambda\Lambda)}\right]_{\Lambda=0}=\left[\frac{\delta\mathcal{L}^{\mathrm{EM}\prime}}{\delta\Lambda}\right]_{\Lambda=0}.\tag{A18}$$

Let us make use of the fact that the current J_λ has definite charge transformation properties. Since $\delta\mathcal{L}/\delta(\partial_\lambda\psi_j)$ transforms as a field with charge $-e_j$, Eqs. (A9) and (A12) tell us that F_j must transform as a field with charge e_j+e_J. Thus,

$$\begin{aligned}F_j[\psi_1\exp(ie_1t),\psi_2\exp(ie_2t),\cdots]\\ =\exp[i(e_j+e_J)t]F_j[\psi_1,\psi_2,\cdots].\end{aligned}\tag{A19}$$

Taking the first derivative with respect to t gives the identity

$$\textstyle\sum_l(\delta F_j/\delta\psi_l)e_l\psi_l=(e_j+e_J)F_j.\tag{A20}$$

Consequently, using $\partial_\sigma F_j=\sum_l(\delta F_j/\delta\psi_l)\partial_\sigma\psi_l$, we obtain

$$\begin{aligned}\partial_\sigma F_j-i(e_j+e_J)A_\sigma F_j&=\textstyle\sum_l(\delta F_j/\delta\psi_l)(\partial_\sigma-ie_lA_\sigma)\psi_l\\ &=\textstyle\sum_l(\delta F_j/\delta\psi_l)\pi_{l\sigma}.\end{aligned}\tag{A21}$$

In other words, $\partial_\sigma F_j-i(e_j+e_J)A_\sigma F_j$ is the same function of $\{\psi\}$, $\{\pi_\sigma\}$ as $\partial_\sigma F_j$ is of $\{\psi\}$, $\{\partial_\sigma\psi\}$. Hence, by comparison of Eq. (A17) with Eq. (A9) it is clear that

$$\begin{aligned}&[\delta\mathcal{L}^{\mathrm{EM}\prime}/\delta(\partial_\lambda\Lambda)]_{\Lambda=0}=J_\lambda[\{\psi\},\{\pi_\sigma\}],\\ &\textstyle\sum_j\{[\delta\mathcal{L}^{\mathrm{EM}\prime}/\delta\psi_j']_{\Lambda=0}F_j+[\delta\mathcal{L}^{\mathrm{EM}\prime}/\delta\pi_{j\sigma}']_{\Lambda=0}\\ &\qquad\times[\partial_\sigma F_j-i(e_j+e_J)A_\sigma F_j]\}=D[\{\psi\},\{\pi_\sigma\}].\end{aligned}\tag{A22}$$

Thus, Eq. (A18) can be rewritten as

$$(\partial_\lambda-ie_JA_\lambda)J_\lambda[\{\psi\},\{\pi_\sigma\}]=D[\{\psi\},\{\pi_\sigma\}].\tag{A23}$$

This completes the proof.

Equation (A23) involves unrenormalized quantities throughout and is *exact*. In the case of PCAC, as considered in the text, $D=C^u\phi_\pi$, where the superscript on C^u denotes that it is unrenormalized. It is trivial to pass from Eq. (A23) to Eq. (33) of the text, which involves only renormalized quantities, if we work to lowest order in the electromagnetic coupling e: All electromagnetic renormalization effects are of second order in e and may be neglected. All strong interaction renormalization effects are contained in the ratio C/C^u, where C is the renormalized constant appearing in Eq. (32) of the text.

Excerpt from S. L. Adler and R. F. Dashen, *Current Algebras and Applications to Particle Physics* (W. A. Benjamin, New York, 1968). Reprinted by permission of Pearson Education, Inc. publishing as Pearson Addison Wesley.

APPENDIX A

We give here a detailed discussion of the chirality method for deriving pion low energy theorems. We first consider a partially-conserved axial-vector current; we

then briefly show why the massless pion, conserved axial-vector current case considered by Nambu and Lurié [Paper 5] gives the same answers. We recall that the "chirality" $\chi(t)$ is defined by

$$\chi(t) = \int d^3x[\mathcal{F}_1^{50}(x) + i\mathcal{F}_2^{50}(x)]$$

$$= F_1^5(t) + iF_2^5(t) \tag{A.1}$$

and from Eqs. (1.93) and (1.100), its time derivative is

$$\frac{d\chi(t)}{dt} = \frac{\sqrt{2}M_N M_\pi^2 g_A}{g_r(0)} \int d^3x \, \Phi_{\pi^+}^\dagger$$

$$= \frac{\sqrt{2}M_N M_\pi^2 g_A}{g_r(0)} \int d^3x \, \Phi_{\pi^-} \tag{A.2}$$

If we define χ^{out} and χ^{in} by

$$\chi^{\text{out}} = \lim_{t\to\infty} \chi(t), \quad \chi^{\text{in}} = \lim_{t\to-\infty} \chi(t) \tag{A.3}$$

integration of Eq. (A.2) from $-\infty$ to ∞ gives

$$\begin{aligned}
\chi^{\text{out}} - \chi^{\text{in}} &= [\sqrt{2}M_N g_A/g_r(0)] \int d^4x \, M_\pi^2 \Phi_{\pi^-} \\
&= [\sqrt{2}M_N g_A/g_r(0)] \\
&\quad \times \int d^4x \, (\Box_x^2 + M_\pi^2)\Phi_{\pi^-} \\
&= [\sqrt{2}M_N g_A/g_r(0)] \int d^4x \, J_{\pi^-}
\end{aligned} \tag{A.4}$$

(We have used the fact that $\int d^4x \, \Box_x^2 \Phi_{\pi^-} = 0$.) Taking the matrix element of Eq. (A.4) between states $\langle\beta(q_2)|$ and $|\alpha(q_1)\rangle$ we get*

---

* Again, we suppress the labels "out" and "in" on the states.

$$\begin{aligned}\langle\beta(q_2)|\chi^{out}-\chi^{in}|\alpha(q_1)\rangle &= [\sqrt{2}M_N g_A/g_r(0)] \\ &\quad\times\int d^4x\, e^{i(q_2-q_1)\cdot x} \\ &\quad\times\langle\beta(q_2)|J_{\pi^-}(0)|\alpha(q_1)\rangle \\ &= [\sqrt{2}M_N g_A/g_r(0)] \\ &\quad\times(2\pi)^4\,\delta^4(q_2-q_1) \\ &\quad\times\langle\beta(q_2)|J_{\pi^-}(0)|\alpha(q_1)\rangle\end{aligned} \tag{A.5}$$

Clearly, the right hand side of Eq. (A.5) is the matrix element for the emission of a pion of zero four-momentum in the process $\alpha\to\beta$. We will now show that the left hand side of Eq. (A.5) can be expressed in terms of the S-matrix element $\langle\beta(q_2)|\alpha(q_1)\rangle$ describing the process in the absence of the soft pion.

To see this we must determine the effect of χ^{out} in Eq. (A.5). (The discussion for χ^{in} will be similar.) We know that (i) χ^{out} is the space integral of a local operator, (ii) χ^{out} is time independent [oscillatory terms in $\chi(t)$ vanish in the limit as $t\to\infty$]. Let us suppose that χ^{out} can be written in the form

$$\chi^{out} = \int d^3x\,\mathcal{P}[\{\Phi^{out}\}] \tag{A.6}$$

where $\mathcal{P}$ is a (possibly infinite) polynomial in the "out" fields of all particles present. Consider a term in $\mathcal{P}$ coming from the product of N "out" fields. It has the time dependence $\exp(-i\Omega t)$, with

$$\Omega = \sum_{j=1}^{N}\epsilon_j(\mathbf{p}_j^2+M_j^2)^{1/2} \tag{A.7}$$

with $\epsilon_j=\pm1$ and with the momenta $\mathbf{p}_j$ constrained by the $\mathbf{x}$ integration to satisfy

$$\sum_{j=1}^{N} p_j = 0 \tag{A.8}$$

If no zero mass particles are present, it is easy to show that the constraint of Eq. (A.8) implies that Ω vanishes identically if and only if $N = 2$, $M_1 = M_2$ and $\epsilon_1 = -\epsilon_2$. In other words, time independence requires that χ^{out} be a bilinear expression in the "out" fields and their adjoints of the form

$$\chi^{out} = \sum_j \int d^3x \, \Phi_j^{out(+)}(x)^\dagger O_j^{(+)} \Phi_j^{out(+)}(x) + \sum_j \int d^3x \, \Phi_j^{out(-)}(x)^\dagger O_j^{(-)} \Phi_j^{out(-)}(x) \tag{A.9}$$

The superscripts (+), (-) indicate respectively the positive, negative frequency parts of the "out" fields; the sum extends over all stable particles.

The matrices $O_j^{(\pm)}$ are determined by (iii) Lorentz covariance and parity (χ^{out} is an axial-vector charge), (iv) isotopic spin (χ^{out} is the isospin raising member of an isotopic triplet) and (v) the asymptotic definition of χ^{out} as $\lim_{t\to\infty} \chi(t)$. For example, (iii) and (iv) require that the nucleon term in χ^{out} have the form

$$g^{(+)} \int d^3x \, \overline{\psi}_p^{out(+)}(x) \gamma^0 \gamma_5 \psi_n^{out(+)}(x) = g^{(+)} \int d^3x \, \overline{\psi}_N^{out(+)}(x) \tau_+ \gamma^0 \gamma_5 \psi_N^{out(+)}(x) \tag{A.10}$$

To determine $g^{(+)}$ we use (v), which states that

$$\lim_{t\to\infty} \langle p(q_2) | \chi(t) | n(q_1) \rangle = \langle p(q_2) | \chi^{out} | n(q_1) \rangle \tag{A.11}$$

The matrix element $\langle p(q_2) \mid \chi(t) \mid n(q_1) \rangle$ is actually time independent, because the n and p have equal mass; using Eq. (1.92) to evaluate it gives for the left hand side of Eq. (A.11)

$$\left(\frac{M_N^2}{q_2^0 q_1^0}\right)^{1/2} \bar{u}_p(q_2) g_A \gamma^0 \gamma_5 u_n(q_1)(2\pi)^3 \delta^3(\mathbf{q}_2 - \mathbf{q}_1) \quad (A.12)$$

while Eq. (A.10) implies that the right hand side of Eq. (A.11) is

$$\left(\frac{M_N^2}{q_2^0 q_1^0}\right)^{1/2} \bar{u}_p(q_2) g^{(+)} \gamma^0 \gamma_5 u_n(q_1)(2\pi)^3 \delta^3(\mathbf{q}_2 - \mathbf{q}_1) \quad (A.13)$$

Hence $g^{(+)} = g_A$, and

$$\chi^{\text{out}} = g_A \int d^3x\, \bar{\psi}_N^{\text{out}(+)}(x)\, \tau_+ \gamma^0 \gamma_5 \psi_N^{\text{out}(+)}(x)$$
$$+ \text{ negative frequency part}$$
$$+ \text{ terms for other particles} \quad (A.14)$$

Clearly, the effect of χ^{out} on an outgoing particle is to give back the same particle, with the same momentum, but with changed spin and isotopic spin. The expression for χ^{in} is obtained by replacing all labels "out" in Eq. (A.14) by "in."

Using Eq. (A.14), we can evaluate the matrix element $\langle \beta(q_2) \mid \chi^{\text{out}} \mid \alpha(q_1) \rangle$ in terms of $\langle \beta(q_2) \mid \alpha(q_1) \rangle$. Eq. (A.14) instructs us to form a sum over "insertions" in the lines of all outgoing particles in $\langle \beta(q_2) \mid$. If we write $\langle \beta \mid = \langle \zeta, \ldots \mid$ the contribution of the particle ζ is

$$
\begin{aligned}
&[\langle\beta\,|\,\chi^{\text{out}}\,|\,\alpha\rangle]_{\zeta} \\
&\quad = [\langle\zeta,\ldots|\,\chi^{\text{out}}\,|\,\alpha\rangle]_{\zeta} \\
&\quad = \sum_{\substack{\text{spin, isospin}\\ \text{of } \zeta'}} \int \frac{d^3 q_{\zeta'}}{(2\pi)^3} \\
&\quad \times \langle\zeta\,|\,\chi^{\text{out}}\,|\,\zeta'\rangle\langle\zeta',\ldots|\,\alpha\rangle
\end{aligned}
\tag{A.15}
$$

and the total matrix element is

$$
\langle\beta\,|\,\chi^{\text{out}}\,|\,\alpha\rangle = \sum_{\zeta\in\beta} [\langle\beta\,|\,\chi^{\text{out}}\,|\,\alpha\rangle]_{\zeta}
\tag{A.16}
$$

Let us illustrate with the case of a final nucleon line. We write

$$
\begin{aligned}
\langle\beta(q_2)\,|\,\alpha(q_1)\rangle &= \delta_{\beta,\alpha} + (2\pi)^4 i\ \delta^4(q_2 - q_1)T(\alpha\to\beta) \\
T(\alpha\to\beta) &= \left(\frac{M_N}{q_N^0}\right)^{1/2} i\,\bar{u}_N(q_N)\mathfrak{M}
\end{aligned}
\tag{A.17}
$$

Then the contribution of the final nucleon N to

$$
\langle\beta(q_2)\,|\,\chi^{\text{out}}\,|\,\alpha(q_1)\rangle
$$

is*

$$
\begin{aligned}
&(2\pi)^4 i\,\delta^4(q_2 - q_1)\left(\frac{M_N}{q_N^0}\right)^{1/2} i\,\bar{u}_N(q_N)g_A\,\tau_+\gamma^0\gamma_5 \\
&\times \left(\frac{\not{q}_N + M_N}{2q_N^0}\right)\mathfrak{M}
\end{aligned}
\tag{A.18}
$$

*To compare Eq. (A.18) with Paper 6, which uses the Pauli metric, replace $\not{q}_N$ in Eq. (A.18) by $-i\not{q}_N$. See the introductory remarks about metrics.

that is, the effect of χ^{out} is to cause the insertion

$$\bar{u}_N(q_N) \to \bar{u}_N(q_N) g_A \tau_+ \gamma^0 \gamma_5 \left(\frac{\not{q}_N + M_N}{2q_N^0}\right) \qquad (A.19)$$

in Eq. (A.17). Comparing with Eq. (A.5), we see that the contribution of the final nucleon N to $\langle\beta\,|\,J_{\pi^-}\,|\,\alpha\rangle$ is

$$-\left(\frac{M_N}{q_N^0}\right)^{1/2} \bar{u}_N(q_N) \frac{g_r(0)}{\sqrt{2}M_N} \tau_+ \gamma^0 \gamma_5 \left(\frac{\not{q}_N + M_N}{2q_N^0}\right) \mathfrak{M} \qquad (A.20)$$

The total expression for $\langle\beta\,|\,J_{\pi^-}\,|\,\alpha\rangle$ will be a sum of terms like Eq. (A.20) for all outgoing particles in $\langle\beta\,|$ and all incoming particles in $|\,\alpha\rangle$.

Now let us briefly discuss why Nambu and Lurié get the same insertion rules in the massless pion case. According to Eq. (A.2), when $M_\pi = 0$ the chirality is conserved,

$$\frac{d\chi_{M_\pi = 0}}{dt} = 0 \qquad (A.21)$$

which implies that $\chi^{out}_{M_\pi = 0} - \chi^{in}_{M_\pi = 0} = 0$. So the term proportional to $\int d^3x\, J_{\pi^-}$ in Eq. (A.4) is no longer present. However, when the pion is massless, the asymptotic chirality χ^{out}, in addition to containing the bilinear terms of Eq. (A.9), will also have a time independent term proportional to $\int d^3x (\partial/\partial x^0)\, \Phi^{out}_{\pi^-}(x)$. This term expresses the fact that emission of a zero four-momentum pion is a *physical* process, not a virtual process, when $M_\pi = 0$. [The $(\partial/\partial x^0)$ is necessary in order to have the space integral of the time component of an axial-vector.] It is easy to verify that

$$\chi^{out}_{M_\pi = 0} = \chi^{out}_{\text{bilinear}} - \frac{\sqrt{2}M_N g_A}{g_r(0)} \times \int d^3x (\partial/\partial x^0)\, \Phi^{out}_{\pi^-}(x) \qquad (A.22)$$

and similarly for $\chi^{\text{in}}_{M_\pi=0}$. Hence $\chi^{\text{out}}_{M_\pi=0} - \chi^{\text{in}}_{M_\pi=0} = 0$ implies that

$$
\begin{aligned}
&\langle \beta(q_2) | \chi^{\text{out}}_{\text{bilinear}} - \chi^{\text{in}}_{\text{bilinear}} | \alpha(q_1) \rangle \\
&\quad = [\sqrt{2} M_N g_A / g_r(0)] \\
&\qquad \times [\langle \beta(q_2) | \int d^3x (\partial/\partial x^0) \Phi^{\text{out}}_{\pi^-}(x) | \alpha(q_1) \rangle \\
&\qquad - \langle \beta(q_2) | \int d^3x (\partial/\partial x^0) \Phi^{\text{in}}_{\pi^-}(x) | \alpha(q_1) \rangle] \qquad \text{(A.23)}
\end{aligned}
$$

The matrix elements on the right hand side of Eq. (A.23) are easily evaluated,

$$
\begin{aligned}
&-\langle \beta(q_2) | \int d^3x (\partial/\partial x^0) \Phi^{\text{in}}_{\pi^-}(x) | \alpha(q_1) \rangle \\
&\quad = \langle \beta(q_2) | \int d^3x (\partial/\partial x^0) \Phi^{\text{out}}_{\pi^-}(x) | \alpha(q_1) \rangle \\
&\quad = \int \frac{d^3q}{(2\pi)^3} [\langle 0 | \int d^3x (\partial/\partial x^0) \Phi^{\text{out}}_{\pi^-}(x) | \pi^-(q) \rangle] \\
&\qquad \times \{ \langle \pi^-(q) \beta(q_2) | \alpha(q_1) \rangle \} \\
&\quad = \int \frac{d^3q}{(2\pi)^3} [(2\pi)^3 \delta^3(\mathbf{q}) (2q^0)^{-1/2} (-iq^0)] \\
&\qquad \times \{ (2\pi)^4 i \, \delta^4(q_2 + q - q_1) (2q^0)^{-1/2} \\
&\qquad \times \langle \beta(q_2) | J_{\pi^-}(0) | \alpha(q_1) \rangle \} \\
&\quad = \tfrac{1}{2} (2\pi)^4 \delta^4(q_2 - q_1) \langle \beta(q_2) | J_{\pi^-}(0) | \alpha(q_1) \rangle \qquad \text{(A.24)}
\end{aligned}
$$

so that Eq. (A.23) becomes

$$
\begin{aligned}
&\langle \beta(q_2) | \chi^{\text{out}}_{\text{bilinear}} - \chi^{\text{in}}_{\text{bilinear}} | \alpha(q_1) \rangle \\
&\quad = [\sqrt{2} M_N g_A / g_r(0)] (2\pi)^4 \\
&\qquad \times \delta^4(q_2 - q_1) \langle \beta(q_2) | J_{\pi^-}(0) | \alpha(q_1) \rangle \qquad \text{(A.25)}
\end{aligned}
$$

This leads to the same insertion rules as were obtained from Eq. (A.5).

VOLUME 14, NUMBER 25 PHYSICAL REVIEW LETTERS 21 JUNE 1965

CALCULATION OF THE AXIAL-VECTOR COUPLING CONSTANT RENORMALIZATION IN β DECAY

Stephen L. Adler*

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received 17 May 1965)

1. Introduction. –We have derived a sum rule expressing the axial-vector coupling-constant renormalization in β decay in terms of off-mass-shell pion-proton total cross sections. This Letter briefly describes the derivation and gives the numerical results, which agree to within five percent with experiment. Full details will be published elsewhere.

The calculation is based on the following assumptions:

(A) The hadronic current responsible for $\Delta S = 0$ leptonic decays is

$$J_\lambda = G_V(J_\lambda^{V1} + iJ_\lambda^{V2} + J_\lambda^{A1} + iJ_\lambda^{A2}), \qquad (1)$$

where G_V is the Fermi coupling constant ($G_V \approx 1.02\times10^{-5}/M_N^2$).[1] Here $J_\lambda^{Va} = :\bar\psi_N\gamma_\lambda\frac{1}{2}\tau^a \times\psi_N + \cdots:$ is the vector current, which we assume to be the same as the isospin current,[2] and $J_\lambda^{Aa} = :\bar\psi_N\gamma_\lambda\gamma_5\frac{1}{2}\tau^a\psi_N + \cdots:$ is the axial-vector current. Since the vector current is conserved, the vector coupling constant is unrenormalized. The renormalized axial-vector coupling constant g_A is defined by

$$\langle N(q)|J_\lambda|N(q)\rangle = (M_N/q_0)G_V\bar u_N(q)(\gamma_\lambda + g_A\gamma_\lambda\gamma_5)\tau^+ u_N(q). \qquad (2)$$

(B) The axial-vector current is partially conserved (PCAC),[3]

$$\partial_\lambda J_\lambda^{Aa} = \frac{-iM_N M_\pi^2 g_A}{g_r K^{NN\pi}(0)}\varphi_\pi^a, \qquad (3)$$

where g_r is the rationalized, renormalized pion-nucleon coupling constant ($g_r^2/4\pi \approx 14.6$), $K^{NN\pi}(0)$ is the pionic form factor of the nucleon, normalized so that $K^{NN\pi}(-M_\pi^2) = 1$, and φ_π^a is the renormalized pion field. According to Eq. (3), the chiralities $\chi^\pm(t) = \int d^3x(J_4^{A1} \pm iJ_4^{A2})$ satisfy

$$\frac{d}{dt}\chi^\pm(t) = \frac{\sqrt{2}M_N M_\pi^2 g_A}{g_r K^{NN\pi}(0)}\int d^3x\,\varphi_{\pi^\pm}. \qquad (4)$$

(C) The axial-vector current satisfies the equal-time commutation relations

$$[J_4^{Aa}(x), J_4^{Ab}(y)]\Big|_{x_0=y_0} = \delta(\vec x - \vec y)i\epsilon^{abc}J_4^{Vc}(x). \qquad (5)$$

This implies that the chiralities satisfy

$$[\chi^+(t), \chi^-(t)] = 2I^3, \qquad (6)$$

where I^3 is the third component of the isotopic spin.

The assumptions (A) are the usual ones for the leptonic decays. The additional hypotheses (B) and (C) are both necessary to obtain the sum rule for g_A. The hypotheses (A)-(C) are mutually consistent, in the sense that there is a renormalizable field theory (the σ model of Gell-Mann and Lévy[4]) in which they are exactly satisfied.

2. Sum rule. – There are two essentially equivalent ways to derive the sum rule for g_A. The first is to use a method proposed recently by Fubini and Furlan.[5] We take the matrix element of Eq. (6) between single proton states $\langle p(q)|$ and $|p(q')\rangle$. The right-hand side gives

$$\langle p(q)|2I^3|p(q')\rangle = (2\pi)^3\delta(\vec{q}-\vec{q}'). \qquad (7)$$

In the matrix element of the commutator we insert a complete set of intermediate states, separating out the one-nucleon term (to which only the neutron contributes):

$$\langle p(q)|[\chi^+(t),\chi^-(t)]|p(q')\rangle$$
$$=\Big\{\sum_{\text{spin}}\int\frac{d^3k}{(2\pi)^3}\langle p(q)|\chi^+(t)|n(k)\rangle\langle n(k)|\chi^-(t)|p(q')\rangle+\sum_{j\neq N}\langle p(q)|\chi^+(t)|j\rangle\langle j|\chi^-(t)|p(q')\rangle\Big\}-(\chi^+\leftrightarrow\chi^-). \qquad (8)$$

The one-neutron term is easily evaluated using Eq. (2), giving

$$(2\pi)^3\delta(\vec{q}-\vec{q}')g_A^{\,2}(1-M_N^{\,2}/q_0^{\,2}). \qquad (9)$$

In the summation over higher intermediate states we make use of Eq. (4), giving

$$\left[\frac{\sqrt{2}M_NM_\pi^{\,2}g_A}{g_rK^{NN\pi}(0)}\right]^2\sum_{j\neq N}\frac{\langle p(q)|\int d^3x\,\varphi_{\pi^+}|j\rangle\langle j|\int d^3x\,\varphi_{\pi^-}|p(q')\rangle}{(q_0-q_{j0})^2}-(\pi^+\leftrightarrow\pi^-). \qquad (10)$$

From Eqs. (9) and (10), we see that there is a family of sum rules, with q_0 as a parameter. In the limit as q_0 approaches infinity, a sum rule for $1-g_A^{\,-2}$ is obtained. Let us assume that the limiting operation can be taken inside the sum over intermediate states in Eq. (10). It is useful to write this sum in the form

$$\sum_{j\neq N}=\int\frac{d^3q_j}{(2\pi)^3}\int_{M_N+M_\pi}^{\infty}dW\sum_{\substack{j\neq N\\ \text{INT}}}\delta(W-M_j), \qquad (11)$$

where q_j is the total momentum and where "INT" denotes the internal variables of the system j. The invariant mass of the system j is M_j. The integrations over x and q_j can be done explicitly, giving a factor $(2\pi)^3\delta(\vec{q}-\vec{q}')$, and constraining $\vec{q}_j$ to be equal to $\vec{q}$. Let us write

$$\langle j|\varphi_{\pi^\pm}(0)|p(q)\rangle=\left(\frac{M_N}{q_0}\frac{M_j}{q_{j0}}\right)^{1/2}F_j^{\,\pm}, \qquad (12)$$

so that $F_j^{\pm}$ is a Lorentz scalar. Then using the facts that $q_{j0}=(q_0^{\,2}+M_j^{\,2}-M_N^{\,2})^{1/2}$ and $(q_0-q_{j0})^{-2}=(q_0+q_{j0})^2/(M_j^{\,2}-M_N^{\,2})^2$, the limit of Eq. (10) becomes

$$\left[\frac{\sqrt{2}M_Ng_A}{g_rK^{NN\pi}(0)}\right]^2(2\pi)^3\delta(\vec{q}-\vec{q}')\int_{M_N+M_\pi}^{\infty}\frac{dW\,M_NW}{(W^2-M_N^{\,2})^2}$$
$$\times\lim_{q_0\to\infty}\left\{\frac{[q_0+(q_0^{\,2}+W^2-M_N^{\,2})^{1/2}]^2}{q_0(q_0^{\,2}+W^2-M_N^{\,2})^{1/2}}\right\}\lim_{q_0\to\infty}\{K^-[W,(q-q_j)^2]-K^+[W,(q-q_j)^2]\}, \qquad (13a)$$

VOLUME 14, NUMBER 25 PHYSICAL REVIEW LETTERS 21 JUNE 1965

with

$$K^{\pm}[W,(q-q_j)^2] = \sum_{\substack{j\neq N\\ \mathrm{INT}}} \delta(W-M_j) M_\pi^4 |F_j^{\pm}|^2. \tag{13b}$$

The limit of the quantity in boldface curly brackets is 4, and the limit of the momentum transfer $(q-q_j)^2 = -[q_0-(q_0^2+M_j^2-M_N^2)^{1/2}]^2$ is 0. It is easy to see that $K^{\pm}(W,0)$ is equal to $[(W^2-M_N^2)/(2\pi M_N)]\times\sigma_0^{\pm}(W)$, where $\sigma_0^{\pm}(W)$ is the total cross section for scattering of a zero-mass $\pi^{\pm}$ on a proton at center-of-mass energy W. Thus we get the simple and exact result

$$1-\frac{1}{g_A^2} = \frac{4M_N^2}{g_r^2 K^{NN\pi}(0)^2}\frac{1}{\pi}\int_{M_N+M_\pi}^{\infty}\frac{WdW}{W^2-M_N^2}[\sigma_0^{+}(W)-\sigma_0^{-}(W)]. \tag{14}$$

Here $\sigma_0^{\pm}(W)$ is the total cross section for scattering of a zero-mass $\pi^{\pm}$ on a proton, at center-of-mass energy W.

The second method of getting the sum rule parallels the derivation from PCAC of a consistency condition on pion-nucleon scattering.[6] Using the identity

$$(d/dt)\langle N|T[\chi^a(t)\chi^b(0)]|N\rangle = \langle N|[\chi^a(t),\chi^b(0)]\delta(t)|N\rangle + \langle N|T[(d/dt)\chi^a(t)\chi^b(0)]|N\rangle, \tag{15}$$

and hypotheses (B) and (C), one obtains the relation

$$1-\frac{1}{g_A^2} = \frac{-2M_N^2}{g_r^2 K^{NN\pi}(0)^2}G(0,0,0,0), \tag{16}$$

where

$$G(\nu,\nu_B,M_\pi^i,M_\pi^f) = \frac{1}{\nu}A^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f) + B^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f). \tag{17}$$

Here A and B are the usual odd-isospin pion-nucleon scattering amplitudes, ν and ν_B are the energy and momentum transfer variables, and M_π^i and M_π^f are, respectively, the masses of the initial and final pion.[7] If $G(\nu,\cdots)$ is assumed to satisfy an unsubtracted dispersion relation in the energy variable ν, Eq. (14) follows from Eq. (17). Thus, the assumption that the limit $(q_0\to\infty)$ may be taken inside the sum over intermediate states in the method of Fubini and Furlan is equivalent to the assumption that $G(\nu,\cdots)$ obeys an unsubtracted dispersion relation. There is evidence that the unsubtracted dispersion relation for $G(\nu,\cdots)$ is valid.[8] Clearly, if a subtraction were required, the sum rule for g_A would be useless.

3. Numerical evaluation. – Because Eq. (14) involves off-mass-shell pion-proton scattering cross sections, a little work is necessary to compare it with experiment. Let us split the right-hand side of Eq. (14) into the sum of three terms:

$$1-\frac{1}{g_A^2} = \frac{4M_N^2}{g_r^2}(R_1+R_2+R_3), \tag{18}$$

with

$$R_1 = -\frac{1}{\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{d}\mathrm{Im}G(\nu,-M_\pi^2/2M_N,M_\pi,M_\pi)$$

$$=\frac{1}{2\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{\nu^2}(\nu^2-M_\pi^2)^{1/2}[\sigma^{+}(\nu)-\sigma^{-}(\nu)], \tag{19a}$$

$$R_2 = \frac{1}{\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{\nu}\mathrm{Im}G(\nu,-M_\pi^2/2M_N,M_\pi,M_\pi) - \frac{1}{\pi}\int_{M_\pi+M_\pi^2/2M_N}^{\infty}\frac{d\nu}{\nu}\mathrm{Im}G(\nu,0,M_\pi,M_\pi), \tag{19b}$$

and

$$R_3 = \frac{1}{\pi}\int_{M_\pi + M_\pi^2/2M_N}^{\infty} \frac{d\nu}{\nu} \mathrm{Im}\left[G(\nu, 0, M_\pi, M_\pi) - \frac{G(\nu,0,0,0)}{K^{NN\pi}(0)^2}\right]. \quad (19c)$$

The dominant term, R_1, involves only the physical pion-proton total cross sections $\sigma^{\pm}$. Numerical evaluation gives[9,10]

$$(4M_N^2/g_r^2)R_1 = 0.254. \quad (20)$$

The term R_2 can be calculated in terms of pion-nucleon scattering phase shifts, giving[11]

$$(4M_N^2/g_r^2)R_2 = 0.155. \quad (21)$$

The term R_3, which describes corrections arising from taking the external pion off the mass shell, cannot be calculated directly from experimental data. In order to estimate this term, we assume that the off-mass-shell partial-wave amplitude $f_{lJI}(W, M_\pi^i, M_\pi^f)$ is given by

$$f_{lJI}(W, M_\pi^i, M_\pi^f) = \frac{f_{lJI}^{\mathrm{B}}(W, M_\pi^i, M_\pi^f)}{f_{lJI}^{\mathrm{B}}(W, M_\pi, M_\pi)} f_{lJI}(W, M_\pi, M_\pi). \quad (22)$$

(Here l = orbital angular momentum, J = total angular momentum, and I = isospin.) The superscript B denotes the Born approximation. Multiplying the physical f_{lJI} by the ratio of the Born approximations gives the off-mass-shell f_{lJI}, the correct threshold behavior, and the correct nearby left-hand singularities. Generalized unitarity implies that the off-mass-shell and the physical partial-wave amplitudes have nearly the same phase; Eq. (22), which gives them identical phases, approximately satisfies this requirement. Numerical evaluation of R_3, using Eq. (22), gives[12]

$$(4M_N^2/g_r^2)R_3 = -0.061. \quad (23)$$

It is possible that this number for R_3 is correct to within 20%.[13]

Combining the three terms of Eq. (18) yields

$$g_A^{\text{theory}} = 1.24. \quad (24)$$

We have not attempted to make a detailed error estimate.[14] The best experimental value for g_A is[15]

$$g_A^{\text{expt}} = 1.18 \pm 0.02. \quad (25)$$

It is interesting that the region around the 600- and 900-MeV pion-nucleon resonances makes an important contribution to the sum rule. If only the contribution of the (3,3) resonance is retained, we get the result $g_A = 1.44$. Thus, the (3,3) resonance does not exhaust the sum rule.

After completing this work, I learned that a similar calculation has been done independently by Weisberger.[16]

---

*Junior Fellow, Society of Fellows.

[1]In the Cabibbo version of universality [N. Cabibbo, Phys. Rev. Letters 10, 531 (1963)], G_V is replaced by $\cos\theta G_V$.

[2]R. P. Feynman and M. Gell-Man, Phys. Rev. 109, 193 (1958).

[3]M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960); Y. Nambu, Phys. Rev. Letters 4, 380 (1960); S. L. Adler, Phys. Rev. 137, B1022 (1965).

[4]Gell-Mann and Lévy, reference 2. The viewpoint that the commutation relations of Eq. (5) may hold exactly is due to Gell-Mann [M. Gell-Mann, Physics 1, 63 (1964)].

[5]S. Fubini and G. Furlan, to be published.

[6]S. L. Adler, reference 3 and to be published; Y. Nambu and D. Lurié, Phys. Rev. 125, 1429 (1962); Y. Namby and E. Shrauner, Phys. Rev. 128, 862 (1962).

[7]In the scattering reaction $\pi(k_1)+p(q_1)\to\pi(k_2)+p(q_2)$, the variables ν, ν_B, M_π^i, and M_π^f are defined by $\nu = -k_1\cdot(q_1+q_2)/2M_N$, $\nu_B = k_1\cdot k_2/2M_N$, $(M_\pi^i)^2 = -k_1^2$, $(M_\pi^f) = -k_2^2$.

[8]First of all, the convergence of the sum rule of Eq. (11) suggests that an unsubtracted dispersion relation is valid. Secondly, B. Amblard et al., Phys. Letters 10, 138 (1964), have shown that the physical forward charge-exchange amplitude $G(\nu, -M_\pi^2/2M_N, M_\pi, M_\pi)$ satisfies an unsubtracted dispersion relation. It would be surprising if this result were changed by the extrapolation of the external pion mass from M_π to 0.

[9]Values of $\sigma^{\pm}$ from 0 to 110 MeV were taken from the smoothed fit of N. P. Klepikov et al., Joint Institute for Nuclear Research Report No. D-584, 1960 (unpublished). From 110 to 4950 MeV we used the tabulation of B. Amblard et al., Phys. Letters 10, 138 (1964) and private communication. Above 4950 MeV, we used the asymptotic formula $\sigma^- - \sigma^+ = 7.73\ \mathrm{mb}\times[k/$

VOLUME 14, NUMBER 25 PHYSICAL REVIEW LETTERS 21 JUNE 1965

$(1 \text{ BeV}/c)]^{-0.7}$ given by G. von Dardel et al., Phys. Rev. Letters 8, 173 (1962). This formula gives a good fit to the experimental data up to 20 BeV/c. The contribution to the sum rule of the region beyond 20 Bev/c is negligible.

[10]For the pion-nucleon coupling constant we used the value $f^2 = g_r^2 M_\pi^2/16\pi M_N^2 = 0.081 \pm 0.002$, quoted by W. S. Woolcock, Proceedings of the Aix-en-Provence Conference on Elementary Particles, 1961 (C.E.N., Saclay, France, 1961), Vol. I, p. 459.

[11]It is convenient to write R_2 as a single integral over pion-nucleon center-of-mass energy W, the integrand of which is the difference of two terms. This integral is sensitive only to low-energy pion-nucleon scattering data, since the two terms in the integrand cancel at high energies. The number quoted in the text was obtained using Roper's $l_m = 3$ phase shifts [L. D. Roper, Phys. Rev. Letters 12, 340 (1964), and private communication], truncating the integral at $W = 11.20\ M_\pi$. The integral is dominated by the (3, 3) resonances: Extending the integral only over the (3, 3) resonance gave $(4M_N^2/g_r^2)R_2 = 0.166$. A third calculation, using simple Breit-Wigner forms for the (3, 3) and the 600- and 900-MeV resonances, and neglecting all other partial waves, gave $(4M_N^2/g_r^2)R_2 = 0.156$. Thus, the value of R_2 is insensitive to "controversial" features of Roper's phases, such as whether the P_{11} wave resonates.

[12]This number was obtained using Roper's phase shifts, truncating the integral at $W = 11.20\ M_\pi$. Extending the integral only over the (3, 3) resonance gave $(4M_N^2/g_r^2)R_3 = -0.066$; evaluating the integral with only Breit-Wigner terms for the low-lying resonances gave $(4M_N^2/g_r^2)R_3 = -0.059$.

[13]To estimate the accuracy of the model, we repeated the calculation of R_3 with the assumption $f_{lJI}(W, 0, 0) = f_{lJI}(W, M_\pi, M_\pi) K^{NN\pi}(0)^2 (W^2 - M_N^2)^{2l} [(W^2 - M_N^2 + M_\pi^2)^2 - 4W^2 M_\pi^2]$, which includes only a threshold correction factor, and a constant factor $K^{NN\pi}(0)^2$ to account for the change in strength of the nearby left-hand singularities. The numerical result for $(4M_N^2/g_r^2)R_3$ was changed by about 20 %, to -0.051.

[14]The variation among different calculations (references 11-13) of R_2 and R_3 gives an idea of the uncertainty in the theoretical result.

[15]C. S. Wu, private communication.

[16]W. I. Weisberger, accompanying Letter [Phys. Rev. Letters 14, 0000 (1965)]. In the numerical evaluation of Weisberger, g_A is calculated from the dominant term R_l, giving $g_A = 1.16$.

CALCULATION OF THE AXIAL-VECTOR COUPLING CONSTANT RENORMALIZATION IN β DECAY. Stephen L. Adler [Phys. Rev. Letters 14, 1051 (1965)].

Please note the following corrections: (1) Delete the redundant sentence immediately following Eq. (14); (2) in Eq. (19a), $d\nu/d$ should be $d\nu/\nu$; (3) in reference 7, $(M_\pi^f) = -k_2^2$ should be $(M_\pi^f)^2 = -k_2^2$; (4) in reference 8, Eq. (11) should be Eq. (14); (5) in reference 13, $[(W^2 - M_N^2 + M_\pi^2)^2 - 4W^2 M_\pi^2]$ should be $[(W^2 - M_N^2 + M_\pi^2)^2 - 4W^2 M_\pi^2]^{-l}$; (6) in reference 16, 0000 should be 1047 and R_l should be R_1; (7) in reference 2, Gell-Man should be Gell-Mann, and in reference 8, Phys. Letters 10 should be Phys. Letter 10.

PHYSICAL REVIEW VOLUME 140, NUMBER 3B 8 NOVEMBER 1965

Sum Rules for the Axial-Vector Coupling-Constant Renormalization in β Decay*

STEPHEN L. ADLER†
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received 7 June 1965)

Starting from the axial-vector current algebra suggested by Gell-Mann and the hypothesis of a partially conserved axial-vector current, we derive a sum rule relating $1-g_A^{-2}$ to off-mass-shell pion-proton total cross sections. Numerical evaluation gives the theoretical prediction $g_A=1.24$, in good agreement with experiment. A similar sum rule for pion-pion scattering can only be satisfied if there is a large low-energy $I=0$, S-wave pion-pion scattering cross section. We suggest tests, in high-energy neutrino reactions, of an algebra suggested by Gell-Mann for the vector and axial-vector current octets.

INTRODUCTION

WITHIN two years after the discovery of parity violation in the weak interactions, the main features of β decay were clarified.[1] It was found that only vector and axial-vector couplings are present. The vector coupling constant was found to be identical with the vector coupling constant in muon decay; the axial-vector coupling constant was found to differ by a factor $g_A\approx 1.2$ from the value expected for a pure $V-A$ interaction. The identity of the vector coupling constants in beta and in muon decay was soon explained by the hypothesis of a conserved vector current (CVC).[2] The value of the axial-vector coupling constant, on the other hand, has remained somewhat of a mystery.[3]

We give, in this paper, a theory of the axial-vector coupling-constant renormalization g_A, based on the axial-vector current algebra suggested by Gell-Mann[4] and on the hypothesis of a partially conserved axial-vector current (PCAC).[5] In Sec. I, we discuss the assumptions made. In Sec. II, we present two derivations of a sum rule relating $1-g_A^{-2}$ to off-mass-shell pion-proton total cross sections. Numerical evaluation of the sum rule, in Sec. III, gives the theoretical prediction $g_A=1.24$. In Sec. IV, we derive a sum rule relating $2g_A^{-2}$ to pion-pion scattering; we find that this sum rule can be satisfied only if there is a large low-energy $I=0$, S-wave pion-pion scattering cross section. In the final section, we propose tests, in high-energy neutrino experiments, of the algebra proposed by Gell-Mann[4] for the vector and the axial-vector current octets. The tests make no assumptions about partial conservation of the currents.

I. ASSUMPTIONS

The sum rules for g_A discussed below are derived from the following assumptions:

(A) The hadronic current responsible for $\Delta S=0$ leptonic decays is

$$J_\lambda=G_V\cos\theta(J_\lambda^{V1}+iJ_\lambda^{V2}+J_\lambda^{A1}+iJ_\lambda^{A2}),\qquad(1)$$

where G_V is the Fermi coupling constant ($G_V\approx 1.02\times10^{-5}/M_N^2$) and $\cos\theta$ is the Cabibbo angle.[6] Here J_λ^{Va} is the vector current, which we assume to be the same as the isospin current, and J_λ^{Aa} is the axial-vector current. In the Fermi theory, we would have had

$$J_\lambda^{Va}=i:\bar\psi_N\gamma_\lambda\tfrac{1}{2}\tau^a\psi_N:,\qquad(2a)$$

$$J_\lambda^{Aa}=i:\bar\psi_N\gamma_\lambda\gamma_5\tfrac{1}{2}\tau^a\psi_N:.\qquad(2b)$$

Actually, we know that mesonic and other terms must be present. Fortunately, in what follows we will not have to assume any specific expressions for J_λ^V and J_λ^A in terms of particle fields.

Since the vector current is conserved, the vector coupling constant is unrenormalized. The renormalized axial-vector coupling constant g_A is defined by

$$\langle N(q)|J_\lambda|N(q)\rangle=(M_N/q_0)G_V\cos\theta\bar u_N(q)\times(\gamma_\lambda+g_A\gamma_\lambda\gamma_5)\tau^+U_N(q).\qquad(3)$$

(B) The axial-vector current is partially conserved (PCAC),

$$\partial_\lambda J_\lambda^{Aa}=\frac{M_NM_\pi^2g_A}{g_rK^{NN\pi}(0)}\phi_\pi^a.\qquad(4)$$

Here g_r is the rationalized, renormalized pion-nucleon coupling constant ($g_r^2/4\pi\approx14.6$), $K^{NN\pi}(0)$ is the pionic form factor of the nucleon, normalized so that $K^{NN\pi}(-M_\pi^2)=1$, and ϕ_π^a is the renormalized pion field.

* An abbreviated version of the calculation of g_A has appeared in *Physical Review Letters* [S. L. Adler, Phys. Rev. Letters 14, 1051 (1965)]. After this calculation was completed, I learned of similar work by Weisberger [W. I. Weisberger, Phys. Rev. Letters 14, 1047 (1965)].

† Junior Fellow, Society of Fellows.

[1] M. Goldhaber, *Proceedings of the 1958 Annual International Conference on High Energy Physics* (CERN, Geneva, 1958), p. 233.

[2] R. P. Feynman and M. Gell-Mann, Phys. Rev. 109, 193 (1958).

[3] Previous papers on the axial-vector coupling constant renormalization include: R. J. Blin-Stoyle, Nuovo Cimento 10, 132 (1958); S. Okubo, *ibid.* 13, 292 (1959); J. Bernstein, M. Gell-Mann, and L. Michel, *ibid.* 16, 560 (1960); A. P. Balachandran, *ibid.* 23, 428 (1962); H. Banerjee, *ibid.* 23, 1168 (1962); V. S. Mathur, R. Nath, and R. P. Saxena, *ibid.* 31, 874 (1964); Y. S. Kim, *ibid.* 36, 523 (1965); Y. Nambu and G. Jona-Lasinio, Phys. Rev. 124, 246 (1961); Nguyen-Van-Hieu, Nucl. Phys. 42, 129 (1963).

[4] M. Gell-Mann, Physics 1, 63 (1964).

[5] M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960); Y. Nambu, Phys. Rev. Letters 4, 380 (1960); S. L. Adler, Phys. Rev. 137, B1022 (1965).

[6] N. Cabibbo, Phys. Rev. Letters 10, 531 (1963).

According to Eq. (4), the chiralities

$$\chi^{\pm}(t)=-i\int d^3x(J_4^{A1}\pm iJ_4^{A2})$$

satisfy

$$\frac{d}{dt}\chi^{\pm}(t)=\frac{\sqrt{2}M_NM_\pi^2g_A}{g_rK^{NN\pi}(0)}\int d^3x\,\phi_{\pi^{\pm}}. \quad (5)$$

(C) The axial-vector current satisfies the equal-time commutation relations

$$[J_4^{Aa}(x),J_4^{Ab}(y)]|_{x_0=y_0}=-\delta(\mathbf{x}-\mathbf{y})\epsilon^{abc}J_4^{Vc}(x). \quad (6)$$

This implies that the chiralities satisfy

$$[\chi^+(t),\chi^-(t)]=2I^3, \quad (7)$$

where I^3 is the third component of the isotopic spin.

The assumptions (A) are the usual ones for the leptonic decays. The vector-axial-vector form of the leptonic weak interactions is, of course, well established.[1] There is also considerable experimental evidence for the hypothesis[2] that the weak vector current J_λ^{Va} is the same as the isospin current.[7]

The hypothesis (B) of a partially conserved axial-vector current (PCAC) was introduced by Gell-Mann and Lévy[5] and by Nambu[6] to explain the successful Goldberger-Treiman relation[8] for charged pion decay. In addition to predicting the Goldberger-Treiman relation, PCAC predicts an experimentally satisfied relation between the pion-nucleon scattering amplitude $A^{\pi N(+)}$ and the pion-nucleon coupling constant g_r.[9]

The commutation relations (C) play an essential role in the calculation. [Note that Eq. (6) is a somewhat stronger assumption than Eq. (7), since even if spatial derivatives of the delta function were present on the right-hand side of Eq. (6), they would integrate to zero in Eq. (7). Only Eq. (7) is actually needed in the derivation below.] The hypothesis that Eq. (6) or Eq. (7) holds exactly is due to Gell-Mann.[4] Gell-Mann and Ne'eman have emphasized[10] that Eq. (7) is the most natural way in which one can make meaningful the idea of universality of strength between the weak couplings of leptons and baryons, without spelling out in detail the construction of J_λ^A from particle fields. Gell-Mann has also pointed out[11] that Eq. (7), by fixing the scale of the axial-vector current relative to the vector current, can, in principle, determine the axial-vector renormalization g_A.

To sum up, Eqs. (1), (3), (5), and (7) are the hypotheses on which our calculation of g_A is based. They are mutually consistent, in the sense that there is a renormalizable field theory (the σ model of Gell-Mann and Lévy[5]), in which they are exactly satisfied.

II. DERIVATIONS OF THE SUM RULE

We give, in this section, two different derivations of a sum rule expressing g_A in terms of off-mass-shell pion-proton total cross sections. A third derivation has been given by Weisberger.[12]

A. Method of Fubini and Furlan

The simplest derivation uses a method proposed recently by Fubini and Furlan.[13] We take the matrix element of Eq. (7) between single-proton states $\langle p(q)|$ and $|p(q')\rangle$. The right-hand side gives

$$\langle p(q)|2I^3|p(q')\rangle=(2\pi)^3\delta(\mathbf{q}-\mathbf{q}'). \quad (8)$$

In the matrix element of the commutator, we insert a complete set of intermediate states, separating out the one-nucleon term (to which only the neutron contributes):

$$\langle p(q)|[\chi^+(t),\chi^-(t)]|p(q')\rangle=\sum_{\text{spin}}\int\frac{d^3k}{(2\pi)^3}\langle p(q)|\chi^+(t)|n(k)\rangle\langle n(k)|\chi^-(t)|p(q')\rangle$$
$$+\sum_{j\neq N}\langle p(q)|\chi^+(t)|j\rangle\langle j|\chi^-(t)|p(q')\rangle-(\chi^+\leftrightarrow\chi^-). \quad (9)$$

The one-neutron term is easily evaluated using Eq. (3), giving

$$\sum_{\text{spin}}\int\frac{d^3k}{(2\pi)^3}\langle p(q)|\chi^+(t)|n(k)\rangle\langle n(k)|\chi^-(t)|p(q')\rangle$$
$$=\int\frac{d^3k}{(2\pi)^3}(2\pi)^3\delta(\mathbf{q}-\mathbf{k})(2\pi)^3\delta(\mathbf{k}-\mathbf{q}')\left(\frac{M_N}{q_0}\frac{M_N}{k_0}\right)g_A^2\bar{u}(q)\gamma_4\gamma_5\left(\frac{k+iM_N}{2iM_N}\right)\gamma_4\gamma_5u(q') \quad (10)$$
$$=(2\pi)^3\delta(\mathbf{q}-\mathbf{q}')g_A^2(1-M_N^2/q_0^2).$$

[7] C. S. Wu, Rev. Mod. Phys. 36, 618 (1964).
[8] M. L. Goldberger and S. B. Treiman, Phys. Rev. 109, 193 (1958).
[9] S. L. Adler, Ref. 5.
[10] M. Gell-Mann and Y. Ne'eman, Ann. Phys. (N. Y.) 30, 360 (1964).
[11] M. Gell-Mann, Phys. Rev. 125, 1067 (1962).
[12] W. I. Weisberger, Phys. Rev. Letters 14, 1047 (1965).
[13] S. Fubini and G. Furlan, Physics 1, 229 (1965).

In the summation over higher intermediate states we make use of Eq. (5), giving

$$\left[\frac{\sqrt{2}M_N M_\pi^2 g_A}{g_r K^{NN\pi}(0)}\right]^2 \sum_{j\neq N} \frac{\langle p(q)|\int d^3x\,\phi_{\pi^+}|j\rangle\langle j|\int d^3x\,\phi_{\pi^-}|p(q')\rangle}{(q_0-q_{j0})^2} - (\pi^+\leftrightarrow\pi^-). \tag{11}$$

From Eqs. (10) and (11), we see that there is a family of sum rules, with q_0 as a parameter. In the limit as q_0 approaches infinity, a sum rule for $1-g_A^{-2}$ is obtained. Let us assume that the limiting operation can be taken *inside* the sum over intermediate states in Eq. (11). It is useful to write this sum in the form

$$\sum_{j\neq N} = \int \frac{d^3q_j}{(2\pi)^3} \int_{M_N+M_\pi}^{\infty} dW \sum_{\substack{j\neq N\\ \mathrm{INT}}} \delta(W-M_j), \tag{12}$$

where $\mathbf{q}_j$ is the total momentum and where "INT" denotes the internal variables of the system j. We have denoted by M_j the invariant mass of the system j. The integrations over x and $\mathbf{q}_j$ can be done explicitly, giving a factor $(2\pi)^3\delta(\mathbf{q}-\mathbf{q}')$ and constraining $\mathbf{q}_j$ to be equal to $\mathbf{q}$. Let us write

$$\langle j|\phi_{\pi^\pm}(0)|p(q)\rangle = ((M_N/q_0)(M_j/q_{j0}))^{1/2} F_j^\pm, \tag{13}$$

so that $F_j^\pm$ is a Lorentz scalar. Then we have for the summation over higher intermediate states,

$$(2\pi)^3\delta(\mathbf{q}-\mathbf{q}')\left[\frac{\sqrt{2}M_N M_\pi^2 g_A}{g_r K^{NN\pi}(0)}\right]^2 \int_{M_N+M_\pi}^{\infty} dW \sum_{\substack{j\neq N\\ \mathrm{INT}}} \delta(W-M_j)(M_N/q_0)(M_j/q_{j0})(q_0-q_{j0})^{-2}[|F_j^-|^2-|F_j^+|^2]. \tag{14}$$

Using the equations

$$q_{j0} = (q_0^2+M_j^2-M_N^2)^{1/2}, \tag{15a}$$

$$(q_0-q_{j0})^{-2} = (q_0+q_{j0})^2/(M_j^2-M_N^2)^2, \tag{15b}$$

the limit as $q_0\to\infty$ of Eq. (14) becomes

$$\left[\frac{\sqrt{2}M_N g_A}{g_r K^{NN\pi}(0)}\right]^2 (2\pi)^3\delta(\mathbf{q}-\mathbf{q}') \int_{M_N+M_\pi}^{\infty} dW \frac{M_N W}{(W^2-M_N^2)^2} \lim_{q_0\to\infty}\left\{\frac{[q_0+(q_0^2+W^2-M_N^2)^{1/2}]^2}{q_0(q_0^2+W^2-M_N^2)^{1/2}}\right\}$$
$$\times \lim_{q_0\to\infty} [K^-[W,(q-q_j)^2]-K^+[W,(q-q_j)^2]], \tag{16}$$

where we have defined $K^\pm[W,(q-q_j)^2]$ by the equation

$$K^\pm[W,(q-q_j)^2] = \sum_{\substack{j\neq N\\ \mathrm{INT}}} \delta(W-M_j) M_\pi^4 |F_j^\pm|^2. \tag{17}$$

Note that $K^\pm$ can only depend on the indicated variables because (i) $K^\pm$ is a Lorentz scalar, and (ii) all internal variables are summed over.[14]

It is now trivial to take the indicated limits. The limit of the quantity in curly brackets is 4, and the limit of the momentum transfer $(q-q_j)^2 = -[q_0-(q_0^2+W^2-M_N^2)^{1/2}]^2$ is 0. Thus we are left with the sum rule

$$1-\frac{1}{g_A^2} = \frac{2M_N^2}{g_r^2 K^{NN\pi}(0)^2} \int_{M_N+M_\pi}^{\infty} \frac{4M_N W dW}{(W^2-M_N^2)^2} [K^+(W,0)-K^-(W,0)]. \tag{18}$$

To complete the derivation, we must express $K^\pm(W,0)$ in terms of pion-proton scattering cross sections. Let $\sigma_0^\pm(W)$ denote the total cross section for scattering of a *zero-mass* $\pi^\pm$ on a proton, at center-of-mass energy W. It is easiest to calculate $\sigma_0^\pm(W)$ in the center-of-mass frame. If we let k and q be, respectively, the four-momenta of

[14] An average over initial proton spin is understood, but is not indicated explicitly.

the initial pion and proton, then we have[14]

$$\sigma_0^{\pm}(W)\cdot\text{flux}=(2\pi)^4\sum_{j\neq N}\frac{|\langle j|J_{\pi^{\pm}}(0)|p(q)\rangle|^2}{2k_0}\delta^4(q_j-q-k)$$
$$=(2\pi)^4\int\frac{d^3q_j}{(2\pi)^3}\sum_{\substack{j\neq N\\ \text{INT}}}\frac{|\langle j|J_{\pi^{\pm}}(0)|p(q)\rangle|^2}{2k_0}\delta^4(q_j-q-k)$$
$$=2\pi\sum_{\substack{j\neq N\\ \text{INT}}}\frac{|\langle j|J_{\pi^{\pm}}(0)|p(q)\rangle|^2}{2k_0}\delta(q_{j0}-q_0-k_0).\tag{19}$$

Keeping in mind the fact that the initial pion has zero mass ($k^2=0$), the following center-of-mass-frame equations may be derived:

$$q_0+k_0=W,\quad q_{j0}\equiv M_j;\tag{20a}$$
$$\text{flux}=|\mathbf{k}|/k_0+|\mathbf{k}|/q_0=W/q_0;\tag{20b}$$
$$k_0=(W^2-M_N^2)/(2W);\tag{20c}$$
$$\langle j|J_{\pi}\pm(0)|p(q)\rangle=M_\pi^2\langle j|\phi_\pi\pm(0)|p(q)\rangle$$
$$=M_\pi^2(M_N/q_0)^{1/2}F_j^{\pm}.\tag{20d}$$

Combining Eqs. (19) and (20) gives

$$\sigma_0^{\pm}(W)=(2\pi M_N/(W^2-M_N^2))\sum_{\substack{j\neq N\\ \text{INT}}}\delta(W-M_j)M_\pi^4|F_j^{\pm}|^2$$
$$=(2\pi M_N/(W^2-M_N^2))K^{\pm}(W,0).\tag{21}$$

Comparing with Eq. (18), we get the simple and exact sum rule

$$1-\frac{1}{g_A^2}=\frac{4M_N^2}{g_r^2K^{NN\pi}(0)^2}\frac{1}{\pi}\int_{M_N+M_\pi}^{\infty}\frac{WdW}{W^2-M_N^2}$$
$$\times[\sigma_0^+(W)-\sigma_0^-(W)].\tag{22}$$

While the derivation just given is straight-forward, it suffers from the defect of requiring an additional assumption: We must assume that the limit $q_0\to\infty$ can be taken inside the sum over interemdiate states in Eq. (11). The next derivation which we give clarifies the meaning of this assumption.

B. "PCAC Consistency Condition" Method

In two previous papers[15] (hereinafter called I and II), we showed that the hypothesis of a partially conserved axial-vector current leads to consistency conditions involving strong-interaction scattering amplitudes. The method used is a general one. Suppose that we have local field operators $j_\lambda(x)$ and $d(x)$ which satisfy the equation

$$\partial_\lambda j_\lambda(x)=d(x).\tag{23}$$

[15] S. L. Adler, Phys. Rev. 137, B1022 (1965), hereinafter called I; S. L. Adler, Phys. Rev. 139, B1638 (1965), hereinafter called II. See also the related papers: Y. Nambu and D. Lurié, *ibid.* 125, 1429 (1962); Y. Nambu and E. Shrauner, *ibid.* 128, 862 (1962).

Let us take the matrix element of this equation between states $\langle\beta(k_F)|$ and $|\alpha(k_I)\rangle$. We get the equation

$$-i(k_F-k_I)_\lambda\langle\beta(k_F)|j_\lambda(0)|\alpha(k_I)\rangle$$
$$=\langle\beta(k_F)|d(0)|\alpha(k_I)\rangle.\tag{24}$$

Let us now consider what happens as $(k_F-k_I)\to0$. In this limit, only those pole terms of $\langle\beta(k_F)|j_\lambda(0)|\alpha(k_I)\rangle$ which behave as $(k_F-k_I)^{-1}$ will contribute to the left-hand side of Eq. (24). It was shown in (II) that these singularities arise only from insertions of the vertex of j_λ on *external* lines of $\langle\beta|\alpha\rangle$. Furthermore, in the limit as $(k_F-k_I)\to0$, these insertions leave the external particles on mass shell. Thus we get a "consistency condition" expressing

$$\lim_{(k_F-k_I)\to0}\langle\beta(k_F)|d(0)|\alpha(k_I)\rangle\tag{25}$$

in terms of the physical matrix element $\langle\beta|\alpha\rangle$. Clearly, the same procedure can be applied to the quantities

$$j(t)=\int d^3x\,j_4(\mathbf{x},t)\quad\text{and}\quad d(t)=\int d^3x\,d(\mathbf{x},t),$$

which satisfy the equation

$$dj(t)/dt=id(t).\tag{26}$$

Of course, the resulting formulas will not be manifestly covariant. What was done in (II) was to study in detail the case when $j(t)$ is simply the chirality $\chi^a(t)$. We will now apply the same method to a somewhat more complicated object,

$$j(x_0)=\int dy_0\,e^{-ik_0y_0}\langle N(q)|T[\chi^a(x_0)\chi^b(y_0)]|N(q)\rangle,\tag{27}$$

in order to rederive the sum rule for g_A.

Let us consider the quantity T defined by

$$T=\int dx_0\,e^{il_0x_0}\int dy_0\,e^{-ik_0y_0}$$
$$\times\langle N(q)|T[\chi^a(x_0)\chi^b(y_0)]|N(q)\rangle$$
$$=\int dx_0\,e^{il_0x_0}j(x_0).\tag{28}$$

Let us also define $P^a(x)$ by the equation

$$\partial_\lambda J_\lambda^{Aa}(x)=P^a(x), \tag{29}$$

so that the chirality $\chi^a(x_0)$ satisfies

$$\frac{d}{dx_0}\chi^a(x_0)=\int d^3x\, P^a(x). \tag{30}$$

We will introduce the assumption that $P^a(x)\propto\phi_\pi^a(x)$ at a later stage of the calculation.

From time-translation invariance, we know that

$$j(x_0)=e^{-ik_0x_0}\times\text{constant}. \tag{31}$$

Consequently,

$$\begin{aligned}-ik_0j(x_0)=\frac{d}{dx_0}j(x_0)&=\int dy_0\, e^{-ik_0y_0}\langle N(q)|\frac{d}{dx_0}T[\chi^a(x_0)\chi^b(y_0)]|N(q)\rangle\\&=e^{-ik_0x_0}\langle N(q)|[\chi^a(x_0),\chi^b(x_0)]|N(q)\rangle+\int dy_0\int d^3x\, e^{-ik_0y_0}\langle N(q)|T[P^a(x)\chi^b(y_0)]|N(q)\rangle.\end{aligned} \tag{32}$$

Since the second term on the right-hand side of Eq. (32) is proportional to $\exp(-ik_0x_0)$, we can rewrite it as

$$\frac{1}{-k_0^2+M_\pi^2}\int dy_0\int d^3x\, e^{-ik_0y_0}(-\Box_x+M_\pi^2)\langle N(q)|T[P^a(x)\chi^b(y_0)]|N(q)\rangle. \tag{33}$$

We have assumed that we can integrate by parts with respect to the *spatial* variables $\mathbf{x}$; this can be justified by the use of wave packets.[16] Combining Eqs. (28), (32), and (33), and then interchanging the order of the integrations over x_0 and y_0, gives

$$\begin{aligned}-ik_0T=&\int dx_0\, e^{i(l_0-k_0)x_0}\langle N(q)|[\chi^a(x_0),\chi^b(x_0)]|N(q)\rangle\\&+\int dx_0\, e^{il_0x_0}\frac{1}{M_\pi^2-k_0^2}\int dy_0\int d^3x\, e^{-ik_0y_0}(-\Box_x+M_\pi^2)\langle N(q)|T[P^a(x)\chi^b(y_0)]|N(q)\rangle\\=&\,2\pi\delta(l_0-k_0)\langle N(q)|[\chi^a(0),\chi^b(0)]|N(q)\rangle+\frac{1}{M_\pi^2-k_0^2}\int dy_0\, e^{-ik_0y_0}j_1(y_0),\end{aligned} \tag{34}$$

with

$$\begin{aligned}j_1(y_0)=&\int d^4x\, e^{il_0x_0}(-\Box_x+M_\pi^2)\langle N(q)|T[P^a(x)\chi^b(y_0)]|N(q)\rangle\\=&\,e^{il_0y_0}\times\text{constant}.\end{aligned} \tag{35}$$

Treating $j_1(y_0)$ in the same manner as we treated $j(x_0)$, we get

$$\begin{aligned}il_0j_1(y_0)=&M_\pi^2\int d^3x\, e^{il_0y_0}\langle N(q)|[\chi^b(y_0),P^a(\mathbf{x},y_0)]|N(q)\rangle\\&+\frac{1}{M_\pi^2-l_0^2}\int d^4x\int d^4y\, e^{il_0x_0}(-\Box_x+M_\pi^2)(-\Box_y+M_\pi^2)\langle N(q)|T[P^a(x)P^b(y)]|N(q)\rangle.\end{aligned} \tag{36}$$

To sum up, we have derived the identity

$$\begin{aligned}-ik_0&\int dx_0\, e^{il_0x_0}\int dy_0\, e^{-ik_0y_0}\langle N(q)|T[\chi^a(x_0)\chi^b(y_0)]|N(q)\rangle\\&=2\pi\delta(l_0-k_0)\left[\langle N(q)|[\chi^a(0),\chi^b(0)]|N(q)\rangle+\left(\frac{M_\pi^2}{M_\pi^2-k_0^2}\right)\frac{1}{il_0}\int d^3x\langle N(q)|[\chi^b(0),P^a(\mathbf{x},0)]|N(q)\rangle\right]\\&+\frac{1}{(M_\pi^2-k_0^2)(M_\pi^2-l_0^2)}\frac{1}{il_0}\int d^4x\int d^4y\, e^{il_0x_0-ik_0y_0}(-\Box_x+M_\pi^2)(-\Box_y+M_\pi^2)\langle N(q)|T[P^a(x)P^b(y)]|N(q)\rangle.\end{aligned} \tag{37}$$

[16] We will never integrate by parts with respect to the time variable.

Since we will obtain the sum rule for g_A from the part of Eq. (37) which is *antisymmetric* in a and b, let us drop all terms which are symmetric. Because $[\chi^a(x_0),\chi^b(x_0)]=i\epsilon^{abc}I^c$, and since $dI^c/dx_0=0$, we have $d[\chi^a(x_0),\chi^b(x_0)]/dx_0=0$. In other words,

$$\int d^3x[P^a(\mathbf{x},x_0),\chi^b(x_0)]=\int d^3x[P^b(\mathbf{x},x_0),\chi^a(x_0)], \tag{38}$$

indicating symmetry under interchange of a and b. Thus we can drop the term proportional to

$$\langle N(q)|[\chi^b(0),P^a(\mathbf{x},0)]|N(q)\rangle.$$

Let us now consider the antisymmetric part of Eq. (37) for small k_0. At the end of the calculation, we will let k_0 approach 0. On the left-hand side, only diagrams with χ^a inserted on the external nucleon lines will make a contribution of zeroth order in k_0, as was shown in (II). This can be seen directly by inserting a complete set of intermediate states in the time-ordered product:

$$\begin{aligned}
&\int dx_0\int dy_0 e^{il_0x_0-ik_0y_0}\langle N(q)|T[\chi^a(x_0)\chi^b(y_0)]|N(q)\rangle\\
&=\int dx_0\int dy_0 e^{il_0x_0-ik_0y_0}\sum_j[\langle N(q)|\chi^a(x_0)|j\rangle\langle j|\chi^b(y_0)|N(q)\rangle\theta(x_0-y_0)\\
&\qquad\qquad+\langle N(q)|\chi^b(y_0)|j\rangle\langle j|\chi^a(x_0)|N(q)\rangle\theta(y_0-x_0)]\\
&=\sum_j[\langle N(q)|J_4^{Aa}(0)|j\rangle\langle j|J_4^{Ab}(0)|N(q)\rangle i(k_0-\Delta_j)^{-1}-\langle N(q)|J_4^{Ab}(0)|j\rangle\langle j|J_4^{Aa}(0)|N(q)\rangle i(k_0+\Delta_j)^{-1}]\\
&\qquad\qquad\times 2\pi\delta(l_0-k_0)(2\pi)^3\delta(0)\delta(\mathbf{q}_j-\mathbf{q}),
\end{aligned}\tag{39}$$

where

$$\Delta_j\equiv(q_0^2+M_j^2-M_N^2)^{1/2}-q_0. \tag{40}$$

Clearly, only the one-nucleon intermediate state ($j=N$, $\Delta_j\equiv0$) gives a singularity behaving as k_0^{-1}. Evaluation of the spin sum, as in Eq. (10), gives, for the left-hand side of Eq. (37),

$$(2\pi)^4\delta(0)\delta(l_0-k_0)g_A^2 i\epsilon^{abc}\langle\tfrac{1}{2}\tau^c\rangle(1-M_N^2/q_0^2)+O(k_0), \tag{41}$$

where $O(k_0)$ indicates terms which vanish as $k_0\to0$.

Let us now evaluate the terms of the right-hand side of Eq. (37). The commutator of the chiralities is easily evaluated, using Eq. (6), giving

$$2\pi\delta(l_0-k_0)\langle N(q)|[\chi^a(0),\chi^b(0)]|N(q)\rangle=(2\pi)^4\delta(0)\delta(l_0-k_0)i\epsilon^{abc}\langle\tfrac{1}{2}\tau^c\rangle. \tag{42}$$

In the last term of Eq. (37), let us introduce the PCAC hypothesis,

$$P^a(x)=\frac{M_NM_\pi^2g_A}{g_rK^{NN\pi}(0)}\phi_\pi^a(x), \tag{43}$$

giving

$$\begin{aligned}
&\left(\frac{M_\pi^2}{M_\pi^2-k_0^2}\right)\left(\frac{M_\pi^2}{M_\pi^2-l_0^2}\right)\left[\frac{M_Ng_A}{g_rK^{NN\pi}(0)}\right]^2\frac{1}{il_0}\int d^4x\int d^4y\, e^{il_0x_0-ik_0y_0}\\
&\qquad\times(-\Box_x+M_\pi^2)(-\Box_y+M_\pi^2)\langle N(q)|T[\phi_\pi^a(x)\phi_\pi^b(y)]|N(q)\rangle.
\end{aligned}\tag{44}$$

Apart from factors, this is just a pion-nucleon scattering amplitude. In fact, the off-mass-shell pion-nucleon scattering amplitudes

$$A^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f)\quad\text{and}\quad B^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f),$$

where M_π^i and M_π^f are, respectively, the masses of the initial and final pion, are defined by[17]

$$\int d^4x \int d^4y\, e^{-il\cdot x} e^{ik\cdot y}(-\Box_x+M_\pi^2)(-\Box_y+M_\pi^2)\langle N(q_2)|T[\phi_\pi^a(x)\phi_\pi^b(y)]|N(q_1)\rangle$$

$$\equiv -i(2\pi)^4\delta(q_1+k-q_2-l)((M_N/q_{10})(M_N/q_{20}))^{1/2}$$

$$\times \bar{u}_N(q_2)\{[A^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f)-ikB^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f)]\tfrac{1}{2}[\tau^a,\tau^b]+\text{isospin symmetric}\}u_N(q_1), \quad (45a)$$

$$\nu_B=k\cdot l/(2M_N), \quad \nu=-k\cdot(q_1+q_2)/(2M_N). \quad (45b)$$

The term B can be separated into pole terms,[17] and a nonpole part which we label $\bar{B}$:

$$B^{\pi N(-)}=(g_r^2/2M_N)K^{NN\pi}[-(M_\pi^i)^2]K^{NN\pi}[-(M_\pi^f)^2]((\nu_B-\nu)^{-1}+(\nu_B+\nu)^{-1})+\bar{B}^{\pi N(-)}. \quad (46)$$

The integral in Eq. (44) is identical with Eq. (45), with

$$l=(0,il_0)=k=(0,ik_0), \quad M_\pi^i=M_\pi^f=k_0; \quad \nu_B=-k_0^2/(2M_N), \quad \nu=q_0k_0/M_N. \quad (47)$$

Combining Eqs. (44), (45), (46), and (47), we find that Eq. (44) becomes

$$(2\pi)^4\delta(0)\delta(l_0-k_0)i\epsilon^{abc}\langle\tfrac{1}{2}\tau^c\rangle\left\{-g_A^2M_N^2/q_0^2-\frac{2M_N^2}{g_r^2K^{NN\pi}(0)^2}\frac{1}{\nu}g_A^2[A^{\pi N(-)}(\nu,0,0,0)+\nu\bar{B}^{\pi N(-)}(\nu,0,0,0)]\right\}+O(k_0^2), \quad (48)$$

with $\nu=q_0k_0/M_N$. The term proportional to $-g_A^2M_N^2/q_0^2$ arises from the Born term in Eq. (46) when the substitutions of Eq. (47) are made, and just cancels the similar term in Eq. (41). Thus, in the limit as $k_0\to 0$, we obtain from Eq. (37) the Lorentz-invariant identity

$$1-\frac{1}{g_A^2}=\frac{-2M_N^2}{g_r^2K^{NN\pi}(0)^2}G(0), \quad (49)$$

where

$$G(\nu)=\nu^{-1}[A^{\pi N(-)}(\nu,0,0,0)+\nu\bar{B}^{\pi N(-)}(\nu,0,0,0)]$$
$$=\nu^{-1}[A^{\pi N(-)}(\nu,0,0,0)+\nu B^{\pi N(-)}(\nu,0,0,0)]. \quad (50)$$

We are able to drop the bar on B because the Born term $(\nu_B-\nu)^{-1}+(\nu_B+\nu)^{-1}$ vanishes identically at $\nu_B=0$.

Equation (49), which follows solely from the assumptions of Sec. I, is our final result. From the crossing and analyticity properties of $A^{\pi N(-)}$ and $B^{\pi N(-)}$, we know that $G(\nu)$ is an even function of ν and is analytic in the ν plane, apart from cuts running from $\pm[M_\pi+M_\pi^2/(2M_N)]$ to $\pm\infty$. Let us assume that $G(\nu)$ satisfies an unsubtracted dispersion relation in the variable ν. Then we may write

$$G(0)=\frac{2}{\pi}\int_{M_\pi+M_\pi^2/(2M_N)}^{\infty}\frac{d\nu}{\nu}\,\mathrm{Im}G(\nu). \quad (51)$$

It is easily verified that

$$\mathrm{Im}G(\nu)=\tfrac{1}{2}(\sigma_0^- - \sigma_0^+). \quad (52)$$

Changing the integration variable from ν to the center-of-mass energy W $[\nu=(W^2-M_N^2)/(2M_N)]$, and combining Eqs. (49), (51), and (52) leads to the sum rule of Eq. (22). *Thus, the assumption that the limit $q_0\to\infty$ may be taken inside the sum over intermediate states in the method of Fubini and Furlan is equivalent to the assumption that $G(\nu)$ obeys an unsubtracted dispersion relation.*

There is evidence that an unsubtracted dispersion relation for $G(\nu)$ is valid. First of all, provided that the Pomeranchuk theorem is valid, the integral in Eq. (22) is convergent. Secondly, Amblard *et al.* and Höhler *et al.* have shown[18] that the forward charge-exchange scattering amplitude

$$A^{\pi N(-)}(\nu,-M_\pi^2/(2M_N),M_\pi,M_\pi)$$
$$+\nu B^{\pi N(-)}(\nu,-M_\pi^2/(2M_N),M_\pi,M_\pi)$$

satisfies an unsubtracted dispersion relation. It would be surprising if this result were changed by the extrapolation of the external pion mass from M_π to 0. Clearly, if a subtraction were required, the sum rule for g_A would be useless.

By writing a dispersion relation for the last term in Eq. (37), *without* assuming the PCAC hypothesis, one gets a sum rule relating $1-g_A^2$ to cross sections measurable in high-energy neutrino experiments. This sum rule is discussed further in Sec. V.

III. NUMERICAL EVALUATION

Because Eq. (22) involves off-mass-shell pion-proton scattering cross sections, a little work is necessary to compare it with experiment. Let us split the right-hand side of Eq. (22) into the sum of three terms:

$$1-g_A^{-2}=(4M_N^2/g_r^2)(R_1+R_2+R_3), \quad (53)$$

[17] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. 106, 1337 (1957).

[18] B. Amblard *et al.*, Phys. Letters 10, 138 (1964); G. Höhler, G. Ebel, and J. Giesecke, Z. Physik 180, 430 (1964).

with

$$R_1=\frac{1}{\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{\nu}\operatorname{Im}G\left(\nu,-\frac{M_\pi^2}{2M_N},M_\pi,M_\pi\right)$$

$$=\frac{1}{2\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{\nu^2}(\nu^2-M_\pi^2)^{1/2}[\sigma^+(\nu)-\sigma^-(\nu)],\qquad(54a)$$

$$R_2=\frac{1}{\pi}\int_{M_\pi}^{\infty}\frac{d\nu}{\nu}\operatorname{Im}G\left(\nu,-\frac{M_\pi^2}{2M_N},M_\pi,M_\pi\right)$$

$$-\frac{1}{\pi}\int_{M_\pi+M_\pi^2/(2M_N)}^{\infty}\frac{d\nu}{\nu}\operatorname{Im}G(\nu,0,M_\pi,M_\pi),\qquad(54b)$$

$$R_3=\frac{1}{\pi}\int_{M_\pi+M_\pi^2/(2M_N)}^{\infty}\frac{d\nu}{\nu}\operatorname{Im}\left[G(\nu,0,M_\pi,M_\pi)-\frac{G(\nu,0,0,0)}{K^{NN\pi}(0)^2}\right],\qquad(54c)$$

$$G(\nu,\nu_B,M_\pi^i,M_\pi^f)\equiv\nu^{-1}[A^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f)+\nu B^{\pi N(-)}(\nu,\nu_B,M_\pi^i,M_\pi^f)].\qquad(54d)$$

There is a definite reason for splitting things up this way. Numerically, we find that $|R_1|>|R_2|>|R_3|$. The dominant term, R_1, involves only the physical pion-proton scattering cross sections $\sigma^\pm$, and thus can be reliably determined. The terms R_2 and R_3 are corrections, which take into account the fact that the sum rule involves the forward charge-exchange scattering amplitude, with both external pions of *zero mass*. The term R_2 can be calculated in terms of pion-nucleon scattering phase shifts. Since it is dominated by the (3,3) resonance, it can be fairly reliably calculated. The term R_3 is less well known, because a model is needed to calculate the off-mass-shell partial wave amplitudes.

We get the following numerical results[19]

$$(4M_N^2/g_r^2)R_1=0.254,$$
$$(4M_N^2/g_r^2)R_2=0.155,\qquad(55)$$
$$(4M_N^2/g_r^2)R_3=-0.061,$$

giving

$$g_A^{\text{theory}}=1.24.\qquad(56)$$

A reasonable error estimate, based upon the variations among the several calculations of R_2 and R_3 discussed below, is ±0.03. The best experimental value is[20]

$$g_A^{\text{expt}}=1.18\pm0.02.\qquad(57)$$

Thus, the sum rule agrees with experiment to within 5%.

It is interesting that the region around the 600- and 900-MeV pion-nucleon resonances makes an important contribution to the sum rule. If only the contribution of the (3,3) resonance is retained, we get the result $g_A=1.44$. In other words, the (3,3) resonance does not exhaust the sum rule.

The remainder of this section deals with the details of the numerical evaluation

A. Calculation of R_1

As stated above, R_1 is calculated directly from the physical pion-proton total cross sections $\sigma^\pm$. Values of $\sigma^\pm$ from 0 to 110 MeV were taken from the smoothed fit of Klepikov *et al.*[21] From 110 to 4950 MeV, we used the tabulation of Amblard *et al.*[22] Above 4950 MeV, we used the asymptotic formula $\sigma^--\sigma^+=7.73$ mb $\times[k/(\text{BeV}/c)]^{-0.7}$ given by von Dardel *et al.*[23] This formula gives a good fit to the experimental data up to 20 BeV/c. Use of this formula beyond 20 BeV/c represents an extrapolation from the present experimental data, and gives

$$\frac{4M_N^2}{g_r^2}\frac{1}{2\pi}\int_{20\text{ BeV}}^{\infty}\frac{d\nu}{\nu^2}(\nu^2-M_\pi^2)^{1/2}(\sigma^+-\sigma^-)=-0.011.\qquad(58)$$

Thus, unless the $[k/(\text{BeV}/c)]^{-0.7}$ asymptotic behavior is very much in error, the region above 20 BeV/c contributes only a few percent of $1-g_A^{-2}$.

B. Calculation of R_2

It is convenient to express R_2 as a single integral over center-of-mass energy W, the integrand of which is the difference of terms referring to $\nu_B=0$ and to $\nu_B=-M_\pi^2/(2M_N)$. The center-of-mass scattering angle ϕ is given by

$$\begin{aligned}y&\equiv\cos\phi=1+M_\pi^2/|\mathbf{k}|^2\quad\text{at}\quad\nu_B=0,\\ y&\equiv\cos\phi=1\quad\text{at}\quad\nu_B=-M_\pi^2/(2M_N),\end{aligned}\qquad(59)$$

where $|\mathbf{k}|$ is the center-of-mass frame pion momentum. Thus we get

$$R_2=-16\int_{M_N+M_\pi}^{\infty}dW\Delta(W),$$

$$\Delta(W)=\frac{W^2}{(W^2-M_N^2)^2}\left[f_1\left(W,1+\frac{M_\pi^2}{|\mathbf{k}|^2}\right)\frac{(W+M_N)^2}{(W+M_N)^2-M_\pi^2}+f_2\left(W,1+\frac{M_\pi^2}{|\mathbf{k}|^2}\right)\frac{(W-M_N)^2}{(W-M_N)^2-M_\pi^2}\right]$$
$$-\frac{W^2}{(W^2-M_N^2-M_\pi^2)^2}[f_1(W,1)+f_2(W,1)],\qquad(60)$$

[19] For the pion-nucleon coupling constant, we used the value $f^2=g_r^2M_\pi^2/(16\pi M_N^2)=0.081\pm0.002$ quoted by W. S. Woolcock, *Proceedings of the Aix-en-Provence International Conference on Elementary Particles* (Centre d'Etudes Nucléaires de Saclay, Seine et Oise, 1961), Vol. I, p. 459.

[20] C. S. Wu (private communication).

[21] N. P. Klepikov *et al.*, Dubna report D-584, 1960 (unpublished).

[22] B. Amblard *et al.*, Ref. 18 and private communication.

[23] G. von Dardel *et al.*, Phys. Rev. Letters **8**, 173 (1962).

with $f_1(W,y)$ and $f_2(W,y)$ the usual center-of-mass pion-nucleon scattering amplitudes. Since f_1 and f_2 are analytic functions of y in an ellipse with foci ± 1 and with semimajor axis $1+2M_\pi^2/|\mathbf{k}|^2$,[24] we can legitimately use partial-wave expansions in calculating f_1 and f_2 in both terms of Eq. (60). The integral is rapidly convergent, since the two terms in $\Delta(W)$ tend to cancel at high energies.

The number $(4M_N^2/g_r^2)R_2=0.155$ quoted in Eq. (55) was obtained by using Roper's $l_m=3$ phase-shift fit,[25] truncating the integral at $W=11.20M_\pi$. (Beyond this energy no phase-shift fit is available.) The integral is dominated by the (3,3) resonance; extending the integral *only* over the (3,3) resonance gave $(4M_N^2/g_r^2)R_2=0.166$. A third calculation, using simple Breit-Wigner forms for the (3,3) and the 600- and 900-MeV resonances, and neglecting all other partial waves, gave $(4M_N^2/g_r^2)R_2=0.156$ when the integral was truncated at $11.20M_\pi$, and $(4M_N^2/g_r^2)R_2=0.145$ when the integral was extended to an upper limit of $W\approx 17M_\pi$. The good agreement of these numbers indicates that R_2 is insensitive to "controversial" features of Roper's phases, such as whether the P_{11} wave resonates.

C. Calculation of R_3

The term R_3, which describes corrections arising from taking the external pion off the mass shell, cannot be calculated directly from experimental data. In order to estimate this term, we must assume a model for the off-mass-shell partial wave amplitude $f_{lJI}(W,M_\pi^i,M_\pi^f)$. (Here l=orbital angular momentum, J=total angular momentum, and I=isospin.)

Actually, in order to evaluate R_3, we only need to know the imaginary part of $f_{lJI}(W,M_\pi^i,M_\pi^f)$. Below the inelastic threshold at $W=M_N+2M_\pi$, generalized unitarity tells us that

$$\operatorname{Im} f_{lJI}(W,M_\pi^i,M_\pi^f) = |\mathbf{k}|\, f_{lJI}(W,M_\pi^i,M_\pi) f_{lJI}(W,M_\pi^f,M_\pi)^*. \quad (61)$$

The intermediate state pion is, of course, on the mass shell. Since only the region around the (3,3) resonance is appreciably affected by taking the external pions off the mass shell, it suffices to study $f_{lJI}(W,M_\pi^i,M_\pi)$ and then to use the elastic unitarity relation of Eq. (61) to get $\operatorname{Im} f_{lJI}(W,M_\pi^i,M_\pi^f)$.

In constructing a model, we use the following information about f_{lJI}:

(i) *Tkreshold behavior*. From kinematic considerations, we know that near the threshold at $W=M_N+M_\pi$, $f_{lJI}(W,M_\pi^i,M_\pi^f)$ will be equal to $(|\mathbf{k}^i||\mathbf{k}^f|)^l$ times slowly varying factors, with

$$|\mathbf{k}^{i,f}|=[(k_0^{i,f})^2-M_\pi^2]^{1/2},$$
$$k_0^{i,f}=[W^2-M_N^2+(M_\pi^{i,f})^2]/(2W). \quad (62)$$

Here $|\mathbf{k}^i|$ and $|\mathbf{k}^f|$ are the center-of-mass momenta of the initial and final pions; when $M_\pi^i=0(M_\pi)$, we denote $|\mathbf{k}^i|$ by $|\mathbf{k}^0|(|\mathbf{k}|)$.

(ii) *Unitarity*. Setting either M_π^i or M_π^f equal to M_π in Eq. (61), we see that $f_{lJI}(W,M_\pi^i,M_\pi)$ has the same phase δ_{lJI} as the true pion-nucleon partial wave amplitude $f_{lJI}(W,M_\pi,M_\pi)$.

(iii) *Left-hand singularities*. Changing the external pion mass changes the left-hand singularities in the partial wave amplitude $f_{lJI}(W,M_\pi^i,M_\pi^f)$. The left-hand singularities closest to the physical region come from the partial wave projection $f_{lJI}^B(W,M_\pi^i,M_\pi^f)$ of the Born approximation (the pole term) in Eq. (46). Reference to Eq. (46) shows that $f_{lJI}^B(W,M_\pi^i,M_\pi^f)$ contains a factor $K^{NN\pi}[-(M_\pi^i)^2]K^{NN\pi}[-(M_\pi^f)^2]$ arising from the change in strength of the coupling of the external pions to nucleons when the external pion mass is changed from the physical value.

A simple model, which takes into account the considerations (i)–(iii), is to take

$$f_{lJI}(W,M_\pi^i,M_\pi) = \frac{f_{lJI}^B(W,M_\pi^i,M_\pi)}{f_{lJI}^B(W,M_\pi,M_\pi)} f_{lJI}(W,M_\pi,M_\pi). \quad (63)$$

Equation (63) gives $f_{lJI}(W,M_\pi^i,M_\pi)$ the same phase as $f_{lJI}(W,M_\pi,M_\pi)$. Multiplying the physical f_{lJI} by the ratio of the Born approximations gives the off-mass-shell f_{lJI} the correct threshold behavior and, approximately, the correct nearby left-hand singularities. A second model is to take

$$f_{lJI}(W,M_\pi^i,M_\pi) \approx (|\mathbf{k}^i|/|\mathbf{k}|)^l K^{NN\pi}[-(M_\pi^i)^2] f_{lJI}(W,M_\pi,M_\pi). \quad (64)$$

Here we have put in only a threshold correction factor and a constant factor $K^{NN\pi}[-(M_\pi^i)^2]$ to account for the change in strength of the nearby left-hand singularities. According to Eq. (61), the first model gives

$$\operatorname{Im} f_{lJI}(W,0,0) = \left[\frac{f_{lJI}^B(W,0,0)}{f_{lJI}^B(W,M_\pi,M_\pi)}\right]^2 \operatorname{Im} f_{lJI}(W,M_\pi,M_\pi), \quad (65)$$

while the second model gives

$$\operatorname{Im} f_{lJI}(W,0,0) = (|\mathbf{k}^0|/|\mathbf{k}|)^{2l} K^{NN\pi}(0)^2 \times \operatorname{Im} f_{lJI}(W,M_\pi,M_\pi). \quad (66)$$

Although Eq. (61) is valid only below the inelastic threshold, we will use Eq. (65) and Eq. (66) above the inelastic threshold as well as below.

Numerical evaluation of Eq. (54c) gives $(4M_N^2/g_r^2)R_3$

[24] This statement assumes the validity of the Mandelstam representation.

[25] L. D. Roper, Phys. Rev. Letters 12, 340 (1964) and private communication.

$=-0.061$ when the model of Eq. (65) is used, and $(4M_N^2/g_r^2)R_3=-0.051$ when we assume Eq. (66). In both cases, Roper's phase-shift fit was used, and the integral was truncated at $W=11.20M_\pi$. Using Eq. (65) integrated only over the (3,3) resonance gave $(4M_N^2/g_r^2)R_3=-0.066$. Evaluating the integral with only Breit-Wigner terms for the low-lying resonances gave similar results. Thus, the quoted value of R_3, while dependent on the model used for going off mass shell, is insensitive to "controversial" features of the phase shifts.

D. Remarks

The terms R_2 and R_3, which come largely from the (3,3) resonance region, give a combined contribution of 0.094, as compared with the contribution of 0.254 coming from R_1. It may at first seem surprising that the effect of R_2 and R_3 is so big, but it is easy to understand this. From Eq. (66), we can see that the main effect of R_2 and R_3 is to multiply $\sigma_{3,3}$, the (3,3) resonance contribution to the integrand of R_1, by a factor

$$|\mathbf{k}^0|^2/|\mathbf{k}|^2. \tag{67}$$

At the peak of the (3,3) resonance, this factor is 1.27. Since the (3,3) contribution to R_1 is 0.43, we expect R_1 to be increased by an amount of order

$$0.27\times 0.43\approx 0.12\,, \tag{68}$$

in rough agreement with the sum of R_2 and R_3.

IV. PION-PION SCATTERING SUM RULE

In Sec. II, we took the matrix element of Eq. (7) between proton states and derived a sum rule relating g_A to pion-proton scattering. Now let us take the matrix element of Eq. (7) between π^+ states. The same manipulations used in the proton case lead to the sum rule

$$\frac{2}{g_A^2}=\frac{4M_N^2}{g_r^2K^{NN\pi}(0)^2}\frac{1}{\pi}\int_{2M_\pi}^{\infty}\frac{WdW}{W^2-M_\pi^2}\times[\sigma_{0\pi}^-(W)-\sigma_{0\pi}^+(W)], \tag{69}$$

where $\sigma_{0\pi}^\pm(W)$ is the total cross section for scattering of a zero mass $\pi^\pm$ on a physical π^+, at center-of-mass energy W. Equation (69) involves g_A^{-2}, rather than $g_A^{-2}-1$, because the one-pion intermediate state contribution vanishes on account of parity. The factor 2 on the left-hand side of Eq. (69) comes from the fact that $\langle\pi^+(q)|2I^3|\pi^+(q')\rangle=2\cdot(2\pi)^3\delta(\mathbf{q}-\mathbf{q}')$.

While, of course, no direct pion-pion scattering data is available, there is enough information on pion-pion resonances to compare Eq. (69) with experiment. First of all, $\sigma_{0\pi}^+(W)$ comes only from $I=2$ scattering. While there are resonances in the low energy $I=0$ and $I=1$ pion-pion scattering, the $I=2$ scattering seems to be small. Thus the right-hand side of Eq. (69) is positive, agreeing in sign with the left-hand side.

Now let us make a quantitative analysis. According to Eq. (57), the left-hand side of Eq. (69) is

$$2/g_A^2=1.43\,. \tag{70}$$

Let us express the right-hand side of Eq. (69) in terms of the variable $s=W^2$, giving

$$\frac{4M_N^2}{g_r^2K^{NN\pi}(0)^2}\frac{1}{2\pi}\int_{4M_\pi^2}^{\infty}\frac{ds}{s-M_\pi^2}[\sigma_{0\pi}^-(s)-\sigma_{0\pi}^+(s)]. \tag{71}$$

As in the proton case, we take account of the fact that the external pion in Eq. (71) is of zero mass by writing

$$\begin{aligned}\sigma_{0\pi}^{l,I}(s)&=K^{NN\pi}(0)^2(|\mathbf{k}^0|/|\mathbf{k}|)^{2l}\sigma_\pi^{l,I}(s)\\&=K^{NN\pi}(0)^2[(s-M_\pi^2)^2/s(s-4M_\pi^2)]^l\sigma_\pi^{l,I}(s)\,,\end{aligned} \tag{72}$$

where $l=$ orbital angular momentum, $I=$ isospin, and $\sigma_\pi^{l,I}(s)$ is the on-mass-shell partial wave cross section. Thus Eq. (71) becomes

$$\frac{4M_N^2}{g_r^2}\frac{1}{2\pi}\int_{4M_\pi^2}^{\infty}\frac{ds}{s-M_\pi^2}\times\left\{\sum_{\substack{l=0\\ l\text{ even}}}^{\infty}\left[\frac{(s-M_\pi^2)^2}{s(s-4M_\pi^2)}\right]^l\tfrac{2}{3}[\sigma_\pi^{l,0}(s)-\sigma_\pi^{l,2}(s)]+\sum_{\substack{l=1\\ l\text{ odd}}}^{\infty}\left[\frac{(s-M_\pi^2)^2}{s(s-4M_\pi^2)}\right]^l\sigma_\pi^{l,1}(s)\right\}. \tag{73}$$

Let us first evaluate the contributions of the two well-established $\pi\pi$ resonances, the $l=I=1$ ρ and the $l=2$, $I=0$ f^0. We parametrize $\sigma_\pi^{1,1}$ and $\sigma_\pi^{2,0}$ in the form[26]

$$\begin{aligned}\sigma_\pi^{1,1}(s)&=\frac{12\pi\gamma_\rho^2\nu^2/(\nu+M_\pi^2)}{(s_\rho-s)^2+\gamma_\rho^2\nu^3/(\nu+M_\pi^2)},\\ \sigma_\pi^{2,0}(s)&=\frac{20\pi\gamma_f^2\nu^4/(\nu+M_\pi^2)}{(s_f-s)^2+\gamma_f^2\nu^5/(\nu+M_\pi^2)},\quad \nu=\tfrac{1}{4}s-M_\pi^2.\end{aligned} \tag{74}$$

The reduced widths γ_ρ^2 and γ_f^2 are related to the experimental full widths at half-maximum Γ_ρ and Γ_f by

$$\gamma_\rho^2=\frac{\nu_\rho+M_\pi^2}{\nu_\rho^3}s_\rho\Gamma_\rho^2,\quad \gamma_f^2=\frac{\nu_f+M_\pi^2}{\nu_f^5}s_f\Gamma_f^2,\quad \nu_{\rho,f}=\tfrac{1}{4}s_{\rho,f}-M_\pi^2. \tag{75}$$

Using the experimental values[27] $s_\rho=29.7M_\pi^2$, $\Gamma_\rho=0.755M_\pi$, $s_f=80.0M_\pi^2$, $\Gamma_f=0.716M_\pi$, we get, for the ρ and f^0 contributions to Eq. (73),

$$\begin{aligned}\rho\text{ contribution}&=0.42\,,\\ f^0\text{ contribution}&=0.11\,.\end{aligned} \tag{76}$$

As a check, we also calculated the ρ and f^0 contribu-

[26] L. A. P. Balázs, Phys. Rev. **129**, 872 (1963).
[27] A. H. Rosenfeld *et al.*, Rev. Mod. Phys. **36**, 977 (1964).

tions in the narrow resonance approximation. This gave 0.35 for the ρ and 0.09 for the f^0 contribution, indicating that resonance shape corrections will not substantially change the numbers of Eq. (76).

The contribution of 0.53 from the ρ and the f^0 is only 37% of the total of 1.43 required by the sum rule. Since the f^0 contribution is so small, and since there seem to be no resonances with $l \geq 3$ in the low-energy region,[27] it should be reasonable to neglect the contribution of terms with $l \geq 3$ in Eq. (73). Rearranging Eq. (69), we get

$$\begin{aligned}\frac{4M_N^2}{g_r^2}\frac{1}{2\pi}\int_{4M_\pi^2}^{\infty}\frac{ds}{s-M_\pi^2}\tfrac{2}{3}\sigma_\pi^{0,0}(s)\\ =\frac{4M_N^2}{g_r^2}\frac{1}{2\pi}\int_{4M_\pi^2}^{\infty}\frac{ds}{s-M_\pi^2}\\ \times\frac{2}{3}\left\{\sigma_\pi^{0,2}(s)+\left[\frac{(s-M_\pi^2)^2}{s(s-4M_\pi^2)}\right]^2\sigma_\pi^{2,2}(s)\right\}\\ +1.43-0.42-0.11\geq 0.9. \quad (77)\end{aligned}$$

Thus, the pion-pion sum rule can be satisfied only if there is a large low-energy $I=0$, S-wave pion-pion scattering cross section.

In order to get an idea of how big the $I=0$, S-wave scattering cross section would have to be in order to satisfy Eq. (77), we evaluated the left-hand side of Eq. (77) using a simple scattering-length parametrization of the $I=0$, S-wave phase shift,[28]

$$\begin{aligned}(\nu/(\nu+M_\pi^2))^{1/2}\cot\delta^{0,0}=1/a_0+H(\nu),\\ H(\nu)=(2/\pi)(\nu/(\nu+M_\pi^2))^{1/2}\\ \times\ln[(\nu/M_\pi^2)^{1/2}+(\nu/M_\pi^2+1)^{1/2}],\end{aligned}\quad (78)$$

which gives

$$\sigma_\pi^{0,0}=\frac{4\pi a_0^2}{a_0^2\nu+(\nu+M_\pi^2)[1+a_0H(\nu)]^2}. \quad (79)$$

We find that Eq. (77) can be satisfied only if $a_0>1.3$ or if $a_0<-0.85$. It is interesting that an $I=0$, S-wave scattering length of the order of a pion Compton wavelength is also suggested by studies of low-energy pion-nucleon scattering[29] and of K_{l4} decays.[30] Needless to say, there is nothing unique about the parametrization of Eq. (78).

V. TESTS OF THE CURRENT ALGEBRA IN HIGH-ENERGY NEUTRINO REACTIONS

The sum rules discussed in the preceding three sections are derived from two principal hypotheses: the axial-vector current commutation relations of Eq. (7) and the partially conserved axial-vector current hypothesis of Eq. (5). In this section, we discuss a sum rule which follows from the axial-vector current algebra alone, regardless of whether PCAC is true. We will also derive sum rules which follow from a proposed algebra of the strangeness-changing currents.

Let us begin by reviewing the theory of leptonic weak interactions of the hadrons. According to Gell-Mann[4] and to Cabibbo,[8] the hadronic weak current is[31]

$$\begin{aligned}J_\lambda^h=(\mathfrak{F}_{1\lambda}+i\mathfrak{F}_{2\lambda}+\mathfrak{F}_{1\lambda}^5+i\mathfrak{F}_{2\lambda}^5)G_V\cos\theta\\ +(\mathfrak{F}_{4\lambda}+i\mathfrak{F}_{5\lambda}+\mathfrak{F}_{4\lambda}^5+i\mathfrak{F}_{5\lambda}^5)G_V\sin\theta. \quad (80)\end{aligned}$$

Here G_V is the Fermi coupling constant and θ is the Cabibbo angle. The vector currents $\mathfrak{F}_{j\lambda}$ and the axial currents $\mathfrak{F}_{j\lambda}^5$ ($j=1,\cdots,8$) each form an SU_3 octet. The SU_3 generalization of the conserved-vector-current (CVC) hypothesis is to assume that the vector currents $\mathfrak{F}_{j\lambda}$ are just the unitary spin currents, with

$$\begin{aligned}\mathfrak{F}_{a\lambda}=I_\lambda^a,\quad a=1,2,3;\\ \mathfrak{F}_{8\lambda}=\tfrac{1}{2}\sqrt{3}Y_\lambda,\end{aligned}\quad (81)$$

where I_λ^a is the isotopic spin current and Y_λ is the hypercharge current. In our new notation, the currents defined in Sec. I are

$$J_\lambda^{Va}=\mathfrak{F}_{a\lambda},\quad J_\lambda^{Aa}=\mathfrak{F}_{a\lambda}^5,\quad a=1,2,3. \quad (82)$$

Let us define vector and axial-vector "charges" F_j and F_j^5 according to

$$F_j=-i\int d^3x\,\mathfrak{F}_{j4};\quad F_j^5=-i\int d^3x\,\mathfrak{F}_{j4}^5. \quad (83)$$

Gell-Mann[4] has postulated that even in the presence of the SU_3 symmetry-breaking interaction, the following commutation relations hold exactly:

$$\begin{aligned}[F_i,F_j]=if_{ijk}F_k,\\ [F_i,F_j^5]=if_{ijk}F_k^5,\\ [F_i^5,F_j^5]=if_{ijk}F_k.\end{aligned}\quad (84)$$

The chirality commutation relation of Eq. (7) is, of course, just a special case of Eq. (84):

$$[F_1^5+iF_2^5,\,F_1^5-iF_2^5]=2F_3. \quad (85)$$

From Eq. (84), we also get the following commutation relation for the "charge" associated with the strangeness changing part of J_λ^h:

$$\begin{aligned}[F_4+iF_5+F_4^5+iF_5^5,\,F_4-iF_5+F_4^5-iF_5^5]\\ =2\sqrt{3}F_8+2F_3+2\sqrt{3}F_8^5+2F_3^5. \quad (86)\end{aligned}$$

Assuming that we can integrate by parts with respect to the spatial variables $\mathbf{x}$, we can express the time derivatives of the "charges" in terms of the divergences of

---

[28] G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).

[29] J. Hamilton, *Strong Interactions and High Energy Physics—Scottish Universities' Summer School, 1963*, edited by R. G. Moorhouse (Plenum Press, New York, 1964).

[30] C. Kacser, P. Singer, and T. Truong, Phys. Rev. **137**, B1605 (1965).

[31] In this section, we use the notation of Ref. 4 for the currents.

Sum Rules for the Axial-Vector Coupling Constant Renormalization in β Decay, STEPHEN L. ADLER [Phys. Rev. **140**, B736 (1965)]. In Eqs. (73) and (77), the coefficient of the isospin-2 cross section $\sigma_\pi^{l,2}$ should be $\frac{5}{3}$ rather than $\frac{2}{3}$. None of the conclusions of Sec. IV is changed. I wish to thank Dr. A. N. Kamal for pointing out this error.

the corresponding currents:

$$\frac{d}{dt}F_j=\int d^3x\,\partial_\lambda \mathfrak{F}_{j\lambda},$$
$$\frac{d}{dt}F_j^5=\int d^3x\,\partial_\lambda \mathfrak{F}_{j\lambda}^5. \tag{87}$$

Let us now derive sum rules which provide tests of the commutation relations of Eq. (85) and Eq. (86), considering first the strangeness-conserving case, Eq. (85). We proceed exactly as in Sec. II, taking the matrix element of Eq. (85) between proton states. The only difference is that we do *not* assume that the divergence $\partial_\lambda \mathfrak{F}_{a\lambda}^5$ is proportional to the pion field. We thus get the sum rule

$$1=g_A^2+\int_{M_N+M_\pi}^{\infty}\frac{4M_N W dW}{(W^2-M_N^2)^2}[N_p^-(W)-N_p^+(W)], \tag{88}$$

with

$$N_p^\pm(W)=\sum_{\substack{j\neq N\\ \text{INT}}}\delta(W-M_j)|\mathfrak{F}_j^\pm|^2|_{(q-q_j)^2=0},$$

$$\langle j|\partial_\lambda\mathfrak{F}_{1\lambda}^5\pm i\partial_\lambda\mathfrak{F}_{2\lambda}^5|p(q)\rangle$$
$$=((M_N/q_0)(M_j/q_{j0}))^{1/2}\mathfrak{F}_j^\pm. \tag{89}$$

In other words, $\mathfrak{F}_j^\pm$ is the matrix element of the divergence of the axial-vector current; the sum rule of Eq. (88) involves this matrix element only at zero four-momentum transfer $(q-q_j)^2$.

The matrix element needed to evaluate the right-hand side of Eq. (88) can be directly measured in high-energy neutrino reactions. Consider the inelastic reaction

$$\nu_l+N\to l+j, \tag{90}$$

with ν_l a neutrino, l a lepton, N a nucleon, and j a system of strongly interacting particles with $M_j\neq M_N$. In a previous paper,[32] we showed that when the lepton emerges parallel to the incident neutrino direction, and when the lepton mass is neglected, the matrix element for Eq. (90) depends only on the *divergences* of the hadronic current. Clearly, under these hypotheses the momentum transfer $(q-q_j)^2$ is zero, so we are measuring just the matrix element needed in Eq. (88). (In the $\Delta S=0$ case, the divergence of the vector current vanishes.) Summing over final states j of strangeness zero ($S=0$) leads to the relations, for forward lepton,

$$\frac{d^2\sigma[\nu+p\to l^-+(S=0)]}{d\Omega_l dE_l}=G_V^2\cos^2\theta f(W)N_p^+(W),$$
$$\frac{d^2\sigma[\bar\nu+p\to l^++(S=0)]}{d\Omega_l dE_l}=G_V^2\cos^2\theta f(W)N_p^-(W), \tag{91}$$

with

$$f(W)=\frac{1}{2\pi^2}\left[\frac{M_N^2+2M_NE-W^2}{W^2-M_N^2}\right]^2. \tag{92}$$

Here E is the incident-neutrino energy, E_l is the final-lepton energy, and Ω_l is the lepton solid angle (all in the laboratory frame, where the initial proton is at rest). In terms of W and E, E_l is given by

$$E_l=(M_N^2+2M_NE-W^2)/(2M_N). \tag{93}$$

We can apply the same method to the commutator of the strangeness-changing currents[33] [Eq. (86)], giving the two sum rules

$$4=\int\frac{4M_NWdW}{(W^2-M_N^2)^2}[S_p^-(W)-S_p^+(W)], \tag{94a}$$

$$2=\int\frac{4M_NWdW}{(W^2-M_N^2)^2}[S_n^-(W)-S_n^+(W)]. \tag{94b}$$

Equation (94a) has discrete contributions at $W=M_\Lambda$, $W=M_\Sigma$ and a continuum from $W=M_\pi+M_\Lambda$ to ∞. Equation (94b) has a discrete contribution at $W=M_\Sigma$ and a continuum from $W=M_\pi+M_\Lambda$ to ∞. The functions $S_{p,n}^\pm$ are measurable in strangeness-changing high-energy neutrino reactions, since for forward lepton,

$$\frac{d^2\sigma[\nu+(p,n)\to l^-+(S=+1)]}{d\Omega_l dE_l}=G_V^2\sin^2\theta f(W)S_{(p,n)}^+(W), \tag{95}$$

$$\frac{d^2\sigma[\bar\nu+(p,n)\to l^++(S=-1)]}{d\Omega_l dE_l}=G_V^2\sin^2\theta f(W)S_{(p,n)}^-(W).$$

Thus, Eqs. (88), (91), (94), and (95) can be used to directly test the algebra proposed by Gell-Mann for the vector and the axial-vector currents.

ACKNOWLEDGMENT

I wish to thank Professor S. B. Treiman for a helpful discussion, and, in particular, for suggesting a study of the pion-pion sum rule.

[32] S. L. Adler, Phys. Rev. 135, B963 (1964).

[33] The nucleon matrix element of the axial-vector terms on the right-hand side of Eq. (86) vanishes when we average over nucleon spin.

Sum Rules for the Axial-Vector Coupling Constant Renormalization in β Decay, STEPHEN L. ADLER [Phys. Rev. **140**, B736 (1965); **149**, 1294(E) (1966)].

1. In the first line of Eq. (62), M_π^2 should read $(M_\pi^{i,f})^2$. In Eq. (65), $f_{IJI}^B(W,0,0)$ should read $f_{IJI}^B(W,0,M_\pi)$. I wish to thank G. E. Brown, A. M. Green, B. H. J. McKellar, and R. Rajaraman for pointing out these errors.

2. A factor of $|\mathbf{k}|/|\mathbf{k}^0|$ was omitted in Eqs. (72), (73), and (77). Equation (72) should read

$$\sigma_{6\pi}^{l,I}(s)=(|\mathbf{k}|/|\mathbf{k}^0|)K^{NN\pi}(0)^2(|\mathbf{k}^0|/|\mathbf{k}|)^{2l}\sigma_\pi^{l,I}(s),$$

and Eqs. (73) and (77) are corrected by making the substitution $ds\to(|\mathbf{k}|/|\mathbf{k}^0|)ds$. Making the correction increases the magnitude of the scattering length a_0 required to saturate the sum rule.

PHYSICAL REVIEW VOLUME 143, NUMBER 4 MARCH 1966

Sum Rules Giving Tests of Local Current Commutation Relations in High-Energy Neutrino Reactions

STEPHEN L. ADLER*
CERN, Geneva, Switzerland and Lyman Laboratory, Harvard University, Cambridge, Massachusetts

(Received 6 October 1965)

We show that the local commutation relations of the vector and the axial-vector current octets can be studied in nonforward lepton-neutrino reactions. We do this by using the commutation relations to derive sum rules, for fixed q^2 (q^2=invariant lepton momentum transfer squared), involving the elastic and the inelastic form factors measured in high-energy neutrino reactions.

1. INTRODUCTION

It has recently been proposed by Gell-Mann[1] that the fourth components of the vector and axial-vector current octets satisfy the local equal-time commutation relations

$$[\mathfrak{F}_{a4}(x),\mathfrak{F}_{b4}(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}(x)\delta(\mathbf{x}-\mathbf{y}), \quad (1a)$$

$$[\mathfrak{F}_{a4}(x),\mathfrak{F}_{b4}{}^5(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}{}^5(x)\delta(\mathbf{x}-\mathbf{y}), \quad (1b)$$

$$[\mathfrak{F}_{a4}{}^5(x),\mathfrak{F}_{b4}{}^5(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}(x)\delta(\mathbf{x}-\mathbf{y}). \quad (1c)$$

Here $\mathfrak{F}_{a\lambda}$ and $\mathfrak{F}_{a\lambda}{}^5$ are, respectively, the octet vector, and axial-vector currents, and a, b, c are unitary spin indices running from 1 to 8. According to Eq. (1), the octet vector and axial-vector charges

$$F_a(t)=-i\int d^3x\ \mathfrak{F}_{a4}(\mathbf{x},t),$$
$$F_a{}^5(t)=-i\int d^3x\ \mathfrak{F}_{a4}{}^5(\mathbf{x},t), \quad (2)$$

satisfy the equal-time commutation relations

$$[F_a(t),F_b(t)]=if_{abc}F_c(t),$$
$$[F_a(t),F_b{}^5(t)]=if_{abc}F_c{}^5(t), \quad (3)$$
$$[F_a{}^5(t),F_b{}^5(t)]=if_{abc}F_c(t).$$

The commutation relations of Eq. (1) are considerably more restrictive than those of Eq. (3), since even if derivatives of the delta function were present on the right-hand side of Eq. (1), Eq. (3) would still be valid. In an earlier paper[2] [hereafter referred to as (I)] we showed that the commutation relations of Eq. (3) can be tested in high-energy inelastic neutrino reactions, in which the lepton (which is regarded as massless) emerges moving parallel to the direction of the incident neutrino. In other words, Eq. (3) may be tested in $q^2=0$ neutrino reactions, where q^2 is the invariant momentum transfer between the neutrino and the outgoing lepton. In this paper we generalize the results of (I), by showing that the *local* commutation relations of Eq. (1) can be tested in $q^2>0$ (nonforward lepton) neutrino reactions. We do this by deriving from Eq. (1) a sum rule, valid for each fixed q^2, involving quantities measurable in high-energy neutrino reactions.

In addition to Eq. (1) for the fourth components of the current octets, let us postulate that the space components of the octets satisfy the local equal-time commutation relations

$$[\mathfrak{F}_{an}(x),\mathfrak{F}_{bm}(y)]|_{x_0=y_0}=\delta_{nm}f_{abc}\mathcal{V}_{c4}{}^1(x)\delta(\mathbf{x}-\mathbf{y})+S_{ab}{}^1, \quad (4a)$$

$$\{[\mathfrak{F}_{an}(x),\mathfrak{F}_{bm}{}^5(y)]|_{x_0=y_0}+[\mathfrak{F}_{an}{}^5(x),\mathfrak{F}_{bm}(y)]|_{x_0=y_0}\}=-2\delta_{nm}f_{abc}\mathfrak{A}_{c4}(x)\delta(\mathbf{x}-\mathbf{y})+S_{ab}{}^2, \quad (4b)$$

$$[\mathfrak{F}_{an}{}^5(x),\mathfrak{F}_{bm}{}^5(y)]|_{x_0=y_0}=\delta_{nm}f_{abc}\mathcal{V}_{c4}{}^2(x)\delta(\mathbf{x}-\mathbf{y})+S_{ab}{}^3. \quad (4c)$$

Here $\mathcal{V}_{c4}{}^1$ and $\mathcal{V}_{c4}{}^2$ are the fourth components of vector-current octets, and $\mathfrak{A}_{c4}$ is similarly the fourth component of an axial-vector octet. The quantities $S_{ab}{}^{1,2,3}$ are symmetric in the unitary spin indices a and b. If the simple quark-model commutation relations proposed by Dashen and Gell-Mann[3] and by Lee[4] are valid, we have

$$\mathcal{V}_{c4}{}^1=\mathcal{V}_{c4}{}^2=\mathfrak{F}_{c4},\quad \mathfrak{A}_{c4}=\mathfrak{F}_{c4}{}^5. \quad (5)$$

However, Eq. (5) is not valid in theories in which meson fields are explicitly included in the currents, whereas, in many of these field theories, Eq. (4) still holds. We will derive sum rules which provide tests of Eq. (4) in $q^2>0$ neutrino reactions.

Each of the sum rules discussed in this paper requires for its derivation, in addition to a local equal-time commutation relation, the assumption that a certain scattering amplitude obeys an unsubtracted dispersion relation in the energy variable, for fixed q^2. *No attempt will be made in this paper to justify the assumption of unsubtracted dispersion relations.* Thus, the statement made in this paper is that if the assumption of unsubtracted dispersion relations is valid, the sum rules derived provide a direct experimental test of local equal-time commutation relations.

In Sec. 2 we state in detail the results of the paper. The next two sections comprise the derivation. In Sec. 3 we analyze the kinematics of high-energy neutrino reactions. In Sec. 4 we derive, from local commutation relations, sum rules which involve the quantities defined in the kinematic analysis of Sec. 3. In an Appendix we give lepton-mass corrections to the results stated in Sec. 2.

* Junior Fellow, Society of Fellows.
[1] M. Gell-Mann, Physics 1, 63 (1964).
[2] S. L. Adler, Phys. Rev. 140, B736 (1965).
[3] R. F. Dashen and M. Gell-Mann, Phys. Letters 17, 142 (1965).
[4] B. W. Lee, Phys. Rev. Letters 14, 676 (1965).

2. RESULTS

We consider the high-energy neutrino reaction

$$\nu+N\to l+\beta\,, \tag{6}$$

where ν is a neutrino, N is a nucleon (neutron or proton), l is an electron or muon, and β is a system of strongly interacting particles. Throughout the text of this paper, we will neglect the final lepton mass, i.e., we take

$$m_l\approx 0. \tag{7}$$

The results stated below are only slightly modified when all lepton-mass terms are included. (See the Appendix.) We define all noninvariant quantities referring to the reaction of Eq. (6) in the *laboratory frame*, in which the nucleon N is at rest:

$$\begin{aligned} E_\nu&=\text{neutrino energy},\\ E_l&=\text{lepton energy},\\ \phi&=\text{lepton-neutrino scattering angle},\\ \Omega_l&=\text{lepton solid angle},\\ k_\nu&=\text{neutrino four-momentum},\\ k_l&=\text{lepton four-momentum},\\ q&=k_\nu-k_l=\text{lepton four-momentum transfer}. \end{aligned} \tag{8a}$$

We denote by W the invariant mass of the system β, by M_N the nucleon mass, and by q^2 the invariant momentum transfer between the leptons:

$$\begin{aligned} q^2&=(k_\nu-k_l)^2=4E_\nu E_l\sin^2(\phi/2)\,,\\ W&=[2M_N(E_\nu-E_l)+M_N{}^2-q^2]^{1/2}. \end{aligned} \tag{8b}$$

We assume that the semileptonic weak interactions are described by the current-current effective Lagrangian density

$$\begin{aligned} \mathcal{L}(x)&=(G/\sqrt{2})j_\lambda(x)J_\lambda(x)+\text{adjoint}\,,\\ G&=1.023\times10^{-5}/M_N{}^2\,,\\ j_\lambda(x)&=\bar\psi_l(x)\gamma_\lambda(1+\gamma_5)\psi_\nu(x)\,,\\ J_\lambda(x)&=(\cos\theta_C)[\mathfrak{F}_{1\lambda}(x)+i\mathfrak{F}_{2\lambda}(x)+\mathfrak{F}_{1\lambda}{}^5(x)+i\mathfrak{F}_{2\lambda}{}^5(x)]\\ &\quad+(\sin\theta_C)[\mathfrak{F}_{4\lambda}(x)+i\mathfrak{F}_{5\lambda}(x)+\mathfrak{F}_{4\lambda}{}^5(x)+i\mathfrak{F}_{5\lambda}{}^5(x)]\,,\\ \theta_C&=\text{Cabibbo angle}. \end{aligned} \tag{9}$$

We define the form factors $F_1{}^V(q^2)$, $F_2{}^V(q^2)$, $g_V(q^2)$, $g_A(q^2)$, and $h_A(q^2)$, which describe elastic neutrino reactions, as follows:

$$\begin{aligned} \langle N(p_2)|\,\mathfrak{F}_{1\lambda}(0)+i\mathfrak{F}_{2\lambda}(0)\,|N(p_1)\rangle&=((M_N/p_{20})(M_N/p_{10}))^{1/2}i\bar u_N(p_2)\tau^+[F_1{}^V(q^2)\gamma_\lambda-F_2{}^V(q^2)\sigma_{\lambda\eta}q_\eta]u_N(p_1)\\ &=((M_N/p_{20})(M_N/p_{10}))^{1/2}i\bar u_N(p_2)\tau^+[g_V(q^2)\gamma_\lambda+iF_2{}^V(q^2)(p_1+p_2)_\lambda]u_N(p_1)\,,\\ q=p_2-p_1,\quad g_V(q^2)&=F_1{}^V(q^2)+2M_NF_2{}^V(q^2)\,,\\ \langle N(p_2)|\,\mathfrak{F}_{1\lambda}{}^5(0)+i\mathfrak{F}_{2\lambda}{}^5(0)\,|N(p_1)\rangle&=((M_N/p_{20})(M_N/p_{10}))^{1/2}i\bar u_N(p_2)\tau^+[g_A(q^2)\gamma_\lambda-ih_A(q^2)q_\lambda]\gamma_5u_N(p_1)\,. \end{aligned} \tag{10}$$

Here τ^+ denotes $\frac{1}{2}(\tau^1+i\tau^2)$, with $\frac{1}{2}\tau^c(c=1,2,3)$ the nucleon isotopic spin matrices.

Finally, we define the diagonal one-nucleon matrix elements of the operators $\mathcal{V}_{c4}{}^{1,2}$ appearing in Eq. (4) as follows

$$\begin{aligned} \langle N(p)|\mathcal{V}_{c4}{}^{1,2}(0)|N(p)\rangle&=iC_I{}^{1,2}\langle\tfrac{1}{2}\tau^c\rangle,\quad c=1,2,3;\\ \langle N(p)|\mathcal{V}_{84}{}^{1,2}(0)|N(p)\rangle&=iC_Y{}^{1,2}. \end{aligned} \tag{11}$$

If the quark-model commutation relations hold, so that Eq. (5) is valid, then

$$C_I{}^{1,2}=1,\quad C_Y{}^{1,2}=\tfrac{1}{2}\sqrt{3}. \tag{12}$$

If the quark-model commutation relations are not valid, the values of $C_I{}^{1,2}$ and $C_Y{}^{1,2}$ are not at present known.

We may now state the results of this paper, as follows.

Strangeness-Conserving Case

The kinematic analysis of Sec. 3 shows that we may write the reaction differential cross section in the form

$$\begin{aligned} d^2\sigma\left(\begin{pmatrix}\nu\\ \bar\nu\end{pmatrix}+p\to\begin{pmatrix}l\\ \bar l\end{pmatrix}+\beta(S=0)\right)\Big/d\Omega_ldE_l&=\frac{G^2\cos^2\theta_C}{(2\pi)^2}\frac{E_l}{E_\nu}\\ &\times[q^2\alpha^{(\pm)}(q^2,W)+2E_\nu E_l\cos^2(\tfrac{1}{2}\phi)\beta^{(\pm)}(q^2,W)\mp(E_\nu+E_l)q^2\gamma^{(\pm)}(q^2,W)]. \end{aligned} \tag{13}$$

By measuring $d^2\sigma/d\Omega_ldE_l$ for various values of the neutrino energy E_ν, the lepton energy E_l, and the lepton-neutrino angle ϕ, we can determine the form factors $\alpha^{(\pm)}$, $\beta^{(\pm)}$, and $\gamma^{(\pm)}$ for all $q^2>0$ and for all W above threshold.

In Sec. 4 we prove that:

(i) the local commutation relations of Eq. (1a) and Eq. (1c) imply

$$2=g_A(q^2)^2+F_1{}^V(q^2)^2+q^2F_2{}^V(q^2)^2+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}dW[\beta^{(-)}(q^2,W)-\beta^{(+)}(q^2,W)]; \tag{14}$$

(ii) the local commutation relations of Eq. (4a) and Eq. (4c) imply

$$C_I^1+C_I^2=(1+q^2/4M_N^2)g_A(q^2)^2+(q^2/4M_N^2)g_V(q^2)^2+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}dW[\alpha^{(-)}(q^2,W)-\alpha^{(+)}(q^2,W)]; \quad (15)$$

(iii) the local commutation relation of Eq. (4b) implies

$$\frac{g_V(q^2)g_A(q^2)}{M_N}=\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}dW[\gamma^{(-)}(q^2,W)-\gamma^{(+)}(q^2,W)]. \quad (16)$$

Strangeness-Changing Case

We write

$$d^2\sigma\left(\begin{pmatrix}\nu\\ \bar\nu\end{pmatrix}+(p,n)\to\begin{pmatrix}l\\ \bar l\end{pmatrix}+\beta\begin{pmatrix}S=1\\ S=-1\end{pmatrix}\right)\Big/ d\Omega_l dE_l=\frac{G^2\sin^2\theta_C}{(2\pi)^2}\frac{E_l}{E_\nu}$$
$$\times[q^2\alpha_{(p,n)}{}^{(\pm)}(q^2,W)+2E_\nu E_l\cos^2(\tfrac{1}{2}\phi)\beta_{(p,n)}{}^{(\pm)}(q^2,W)\mp(E_\nu+E_l)q^2\gamma_{(p,n)}{}^{(\pm)}(q^2,W)]. \quad (17)$$

Then,

(i) the local commutation relations of Eq. (1a) and Eq. (1c) imply

$$(4,2)=\int\frac{W}{M_N}dW[\beta_{(p,n)}{}^{(-)}(q^2,W)-\beta_{(p,n)}{}^{(+)}(q^2,W)]; \quad (18)$$

(ii) the local commutation relations of Eq. (4a) and Eq. (4c) imply

$$[\sqrt{3}(C_Y^1+C_Y^2)+\tfrac{1}{2}(C_I^1+C_I^2),\sqrt{3}(C_Y^1+C_Y^2)-\tfrac{1}{2}(C_I^1+C_I^2)]=\int\frac{W}{M_N}dW[\alpha_{(p,n)}{}^{(-)}(q^2,W)-\alpha_{(p,n)}{}^{(+)}(q^2,W)]; \quad (19)$$

(iii) the local commutation relation of Eq. (4b) implies

$$(0,0)=\int\frac{W}{M_N}dW[\gamma_{(p,n)}{}^{(-)}(q^2,W)-\gamma_{(p,n)}{}^{(+)}(q^2,W)]. \quad (20)$$

The integrals of Eqs. (18)–(20) have discrete contributions at $W=M_\Lambda$ and/or M_Σ and a continuum extending from $W=M_\Lambda+M_\pi$ or from $W=M_\Sigma+M_\pi$ to $W=\infty$. We have not explicitly separated off the discrete contributions to the integrals, as was done in Eqs. (14)–(16) for the strangeness-conserving case. It would, of course, be straightforward to do this.

The sum rules of Eqs. (14)–(16) and (18)–(20) hold for each fixed q^2, provided, as was stated in Sec. 1, that the assumption of an unsubtracted dispersion relation needed to derive each sum rule is valid. When $q^2=0$, Eqs. (41) and (43) of the next section show that

$$\beta(0,W)=(4M_N^2/(W^2-M_N^2)^2)\sum_{\beta,\mathrm{INT}}\sum_s\delta(k_{\beta 0}+E_l-E_\nu-M_N)\{|\langle\beta|\partial_\lambda J_\lambda{}^V|N\rangle|^2+|\langle\beta|\partial_\lambda J_\lambda{}^A|N\rangle|^2\}, \quad (21)$$

where $J_\lambda{}^V$ and $J_\lambda{}^A$ are the vector and axial-vector weak currents appropriate to the $\Delta S=0$ or $|\Delta S|=1$ cases (e.g., $J_\lambda{}^V=\mathfrak{F}_{1\lambda}+i\mathfrak{F}_{2\lambda}$ or $\mathfrak{F}_{4\lambda}+i\mathfrak{F}_{5\lambda}$). Thus, at $q^2=0$ Eqs. (14) and (18) are just the forward lepton sum rules derived in (I).

The sum rule on β has an interesting consequence for the behavior of neutrino cross sections in the limit of very large neutrino energy E_ν. With the aid of Eq. (8), let us write Eqs. (13) and (14) in the form

$$d^2\sigma\left(\begin{pmatrix}\nu\\ \bar\nu\end{pmatrix}+p\to\begin{pmatrix}l\\ \bar l\end{pmatrix}+\beta(S=0)\right)\Big/ d(q^2)dq_0=\frac{G^2\cos^2\theta_C}{4\pi E_\nu^2}[q^2\alpha^{(\pm)}+(2E_\nu^2-2E_\nu q_0-\tfrac{1}{2}q^2)\beta^{(\pm)}\mp(2E_\nu-q_0)q^2\gamma^{(\pm)}], \quad (22)$$

$$2=\int_{(q^2/2M_N)-}^{\infty}dq_0(\beta^{(-)}-\beta^{(+)}). \quad (23)$$

The differential cross section $d\sigma/d(q^2)$ is given by

$$\frac{d\sigma}{dq^2}=\int_{(q^2/2M_N)-}^{E_\nu(1-q^2/4E_\nu^2)}dq_0\frac{d^2\sigma}{d(q^2)dq_0}. \quad (24)$$

The upper limit of integration is fixed by the requirement that $\sin^2(\phi/2)$ lie between 0 and 1. Using Eqs. (22)–(24), it is straightforward to prove the following theorem:

Theorem. Suppose that the integrals

$$\int^{\infty} \frac{dq_0}{q_0^2}(\alpha^{(-)}-\alpha^{(+)}), \quad \int^{\infty} \frac{dq_0}{q_0}(\gamma^{(-)}+\gamma^{(+)}), \quad \int^{\infty} dq_0(\beta^{(-)}-\beta^{(+)}) \tag{25}$$

are convergent. Then

$$\lim_{E_\nu\to\infty} \{d\sigma(\bar{\nu}+p\to l+\beta(S=0))/d(q^2)-d\sigma(\nu+p\to l+\beta(S=0))/d(q^2)\}$$
$$=\frac{G^2\cos^2\theta_C}{2\pi}\int_{(q^2/2M_N)-}^{\infty} dq_0[\beta^{(-)}-\beta^{(+)}]=\frac{G^2\cos^2\theta_C}{\pi}. \tag{26}$$

Similar results hold in the strangeness-changing case. Adding the cross sections for the $\Delta S=0$ and the $|\Delta S|=1$ cases to obtain the total cross section, we find

$$\begin{aligned}&\lim_{E_\nu\to\infty} [d\sigma_T(\bar{\nu}+p)/d(q^2)-d\sigma_T(\nu+p)/d(q^2)]=(G^2/\pi)(\cos^2\theta_C+2\sin^2\theta_C),\\ &\lim_{E_\nu\to\infty} [d\sigma_T(\bar{\nu}+n)/d(q^2)-d\sigma_T(\nu+n)/d(q^2)]=(G^2/\pi)(-\cos^2\theta_C+\sin^2\theta_C).\end{aligned} \tag{27}$$

Equation (27) is the somewhat surprising statement that, in the limit of large neutrino energy, $d\sigma_T(\bar{\nu}+N)/d(q^2)-d\sigma_T(\nu+N)/d(q^2)$ becomes *independent of* q^2. This result is unchanged by the lepton-mass corrections.

3. KINEMATIC ANALYSIS OF HIGH-ENERGY NEUTRINO REACTIONS

In this Section we derive Eq. (13) ,which gives the general form for the neutrino reaction leptonic differential cross section, $d^2\sigma/d\Omega_l dE_l$.[5] In particular, we find explicit expressions for the form factors $\alpha(q^2,W)$, $\beta(q^2,W)$, and $\gamma(q^2,W)$, in terms of matrix elements of the vector and the axial-vector currents.

According to the effective Lagrangian of Eq. (9), the matrix element $\mathfrak{M}$ for the process $\nu+N\to l+\beta$ is given by

$$\mathfrak{M}=gm, \quad m=\bar{u}_l(k_l)\gamma_\lambda(1+\gamma_5)u_\nu(k_\nu)2^{-1/2}\langle\beta^{\text{out}}(k_\beta)|J_\lambda{}^V+J_\lambda{}^A|N(k_N)\rangle. \tag{28}$$

Here $g=(G\cos\theta_C, G\sin\theta_C)$ in $(\Delta S=0, |\Delta S|=1)$ reactions, $J_\lambda{}^V$ and $J_\lambda{}^A$ are the appropriate vector and axial-vector currents, and k_β and k_N are, respectively, the four-momenta of β and of N. In the frame in which the initial nucleon N is at rest, the reaction cross section is given by

$$\sigma=(2\pi)^4\int\frac{d^3k_l}{(2\pi)^3}\int\frac{d^3k_\beta}{(2\pi)^3}\sum_{\beta,\text{INT}}\sum_s \delta(k_\beta+k_l-k_\nu-k_N)\left(\frac{m_l}{E_l}\frac{m_\nu}{E_\nu}\right)g^2\langle|m|^2\rangle. \tag{29}$$

In Eq. (29), $\sum_{\beta,\text{INT}}$ is a sum over the internal variables of the system β, $\sum_s$ is an average over the initial nucleon spin, and $\langle|m|^2\rangle$ is the sum of $|m|^2$ over the lepton spin states. From Eq. (29) we get

$$d^2\sigma/d\Omega_l dE_l=[g^2/(2\pi)^2](E_l/E_\nu)\kappa, \tag{30}$$

with

$$\kappa=\sum_{\beta,\text{INT}}\sum_s \delta(k_{\beta 0}+E_l-E_\nu-M_N)m_l m_\nu\langle|m|^2\rangle|_{\mathbf{k}_\beta=\mathbf{q}}. \tag{31}$$

Let us now study the quantity κ. We introduce the abbreviated notation

$$\begin{aligned}&e_\lambda=2^{-1/2}\bar{u}_l(k_l)\gamma_\lambda(1+\gamma_5)u_\nu(k_\nu),\\ &V_\lambda{}^\beta=\langle\beta^{\text{out}}[\mathbf{q}, i(q_0+M_N)]|J_\lambda{}^V|N(0,iM_N)\rangle,\\ &A_\lambda{}^\beta=\langle\beta^{\text{out}}[\mathbf{q}, i(q_0+M_N)]|J_\lambda{}^A|N(0,iM_N)\rangle,\\ &\Sigma_\beta=\sum_{\beta,\text{INT}}\sum_s \delta(k_{\beta 0}+E_l-E_\nu-M_N).\end{aligned} \tag{32}$$

Let us further denote by $V_D{}^\beta$ and by $A_D{}^\beta$ the matrix elements of the *divergences* of the vector and the axial-vector currents,

$$\begin{aligned}&V_D{}^\beta=-iq_\lambda V_\lambda{}^\beta=\langle\beta^{\text{out}}[\mathbf{q}, i(q_0+M_N)]|\partial_\lambda J_\lambda{}^V|N(0,iM_N)\rangle,\\ &A_D{}^\beta=-iq_\lambda A_\lambda{}^\beta=\langle\beta^{\text{out}}[\mathbf{q}, i(q_0+M_N)]|\partial_\lambda J_\lambda{}^A|N(0,iM_N)\rangle.\end{aligned} \tag{33}$$

[5] Locality theorems of this type are, of course, well known. See, for example, T. D. Lee and C. N. Yang, Phys. Rev. **126**, 2239 (1962); A. Pais, Phys. Rev. Letters **9**, 117 (1962).

Since the final lepton mass is neglected, we have

$$q_\lambda e_\lambda = 0. \tag{34}$$

Using Eqs. (33) and (34), we may write

$$m = e_\lambda(V_\lambda{}^\beta + A_\lambda{}^\beta) = e_n(\delta_{nk} - q_n q_k/q_0{}^2)(V_k{}^\beta + A_k{}^\beta) + i(\mathbf{q}\cdot\mathbf{e}/q_0{}^2)(V_D{}^\beta + A_D{}^\beta), \tag{35}$$

where the repeated indices n and k are summed from 1 to 3. Defining t_{nm} by

$$t_{nm} = \langle e_n e_m{}^* \rangle m_\nu m_l = (k_\nu)_n(k_l)_m + (k_l)_n(k_\nu)_m - k_\nu\cdot k_l \delta_{nm} + \epsilon_{nm\xi\eta}(k_\nu)_\xi(k_l)_\eta, \tag{36}$$

we find that

$$\begin{aligned}
\kappa &= \textstyle\sum_\beta m_l m_\nu \langle |m|^2\rangle |_{\mathbf{k}_\beta=\mathbf{q}} \\
&= t_{nm}(\delta_{mj} - q_m q_j/q_0{}^2)(\delta_{nk} - q_n q_k/q_0{}^2)\{\textstyle\sum_\beta (V_j{}^\beta)^* V_k{}^\beta + \sum_\beta (A_j{}^\beta)^* A_k{}^\beta + \sum_\beta[(A_j{}^\beta)^* V_k{}^\beta + (V_j{}^\beta)^* A_k{}^\beta]\} \\
&\quad + t_{nm}(q_n q_m/q_0{}^4)\{\textstyle\sum_\beta |V_D{}^\beta|^2 + \sum_\beta |A_D{}^\beta|^2 + \sum_\beta[(A_D{}^\beta)^* V_D{}^\beta + (V_D{}^\beta)^* A_D{}^\beta]\} \\
&\quad + t_{nm}(\delta_{nk} - q_n q_k/q_0{}^2)(iq_m/q_0{}^2)\{\textstyle\sum_\beta[(V_k{}^\beta)^* V_D{}^\beta - (V_D{}^\beta)^* V_k{}^\beta] + \sum_\beta[(A_k{}^\beta)^* A_D{}^\beta - (A_D{}^\beta)^* A_k{}^\beta] \\
&\qquad + \textstyle\sum_\beta[(V_k{}^\beta)^* A_D{}^\beta - (A_D{}^\beta)^* V_k{}^\beta] + \sum_\beta[(A_k{}^\beta)^* V_D{}^\beta - (V_D{}^\beta)^* A_k{}^\beta]\}.
\end{aligned} \tag{37}$$

The next step is to use the transformation properties of the currents under time reversal and parity to determine the form of the various $\sum_\beta$ terms in Eq. (37). Denoting by T and by P the time-reversal and parity operators, respectively, we have

$$TJ_k{}^V(0)T^{-1} = -J_k{}^V(0), \quad TJ_k{}^A(0)T^{-1} = -J_k{}^A(0), \quad PJ_k{}^V(0)P^{-1} = -J_k{}^V(0), \quad PJ_k{}^A(0)P^{-1} = J_k{}^A(0), \tag{38}$$

and similarly for the divergences of the currents. Under the assumption that the "in" and "out" states of definite total energy *each* form a complete basis for states of that energy, we have

$$\sum_{\beta,\mathrm{INT}} \delta(k_{\beta 0} + E_l - E_\nu - M_N)|\beta^{\mathrm{out}}(k_\beta)\rangle\langle\beta^{\mathrm{out}}(k_\beta)| = \sum_{\beta,\mathrm{INT}} \delta(k_{\beta 0} + E_l - E_\nu - M_N)|PT\beta^{\mathrm{out}}(k_\beta)\rangle\langle PT\beta^{\mathrm{out}}(k_\beta)|, \tag{39a}$$

and

$$\sum_s |N(k_N)\rangle\langle N(k_N)| = \sum_s |PTN(k_N)\rangle\langle PTN(k_N)|. \tag{39b}$$

Using Eqs. (38) and (39) we find that

$$\begin{aligned}
\textstyle\sum_\beta V_k{}^\beta(V_j{}^\beta)^* &= \sum_{\beta,\mathrm{INT}} \sum_s \delta(k_{\beta 0} + E_l - E_\nu - M_N)\langle\beta^{\mathrm{out}}|J_k{}^V|N\rangle\langle\beta^{\mathrm{out}}|J_j{}^V|N\rangle^* \\
&= \sum_{\beta,\mathrm{INT}} \sum_s \delta(k_{\beta 0} + E_l - E_\nu - M_N)\langle PT\beta^{\mathrm{out}}|J_k{}^V|PTN\rangle^*\langle PT\beta^{\mathrm{out}}|J_j{}^V|PTN\rangle \\
&= \textstyle\sum_\beta V_j{}^\beta(V_k{}^\beta)^* = [\sum_\beta V_k{}^\beta(V_j{}^\beta)^*]^*.
\end{aligned} \tag{40}$$

Thus, the tensor $\sum_\beta V_k{}^\beta(V_j{}^\beta)^*$ is *real*, and hence *symmetric*. Using P alone shows that this tensor is an *even* function of $\mathbf{q}$. A similar analysis can be carried through for each of the $\sum_\beta$ terms in Eq. (37), with the following results:

(i) $\sum_\beta V_k{}^\beta(V_j{}^\beta)^*$ and $\sum_\beta A_k{}^\beta(A_j{}^\beta)^*$ are real symmetric tensors (even under $\mathbf{q}\to-\mathbf{q}$);

(ii) $\sum_\beta[V_k{}^\beta(A_j{}^\beta)^* + A_k{}^\beta(V_j{}^\beta)^*]$ is an imaginary, antisymmetric pseudotensor (odd under $\mathbf{q}\to-\mathbf{q}$);

(iii) $\sum_\beta|V_D{}^\beta|^2$ and $\sum_\beta|A_D{}^\beta|^2$ are real scalars;

(iv) $\sum_\beta[V_D{}^\beta(A_D{}^\beta)^* + A_D{}^\beta(V_D{}^\beta)^*]$ is an imaginary pseudoscalar;

(v) $\sum_\beta[V_k{}^\beta(V_D{}^\beta)^* - (V_k{}^\beta)^* V_D{}^\beta]$ and $\sum_\beta[A_k{}^\beta(A_D{}^\beta)^* - (A_k{}^\beta)^* A_D{}^\beta]$ are imaginary vectors;

(vi) $\sum_\beta[V_k{}^\beta(A_D{}^\beta)^* - (V_k{}^\beta)^* A_D{}^\beta]$ and $\sum_\beta[A_k{}^\beta(V_D{}^\beta)^* - (A_k{}^\beta)^* V_D{}^\beta]$ are imaginary pseudovectors.

All of these quantities must be formed from the one vector available, $\mathbf{q}$. Thus the only possible tensors are δ_{kj} and $q_k q_j$ and the only pseudotensor is $\epsilon_{kjn}q_n$. No pseudovectors or pseudoscalars can be formed. Consequently, the most general from of the quantities appearing in Eq. (37) is

$$\begin{aligned}
&\textstyle\sum_\beta (V_j{}^\beta)^* V_k{}^\beta = \delta_{jk} V_1(q^2,W) + q_j q_k V_2(q^2,W), \\
&\textstyle\sum_\beta (A_j{}^\beta)^* A_k{}^\beta = \delta_{jk} A_1(q^2,W) + q_j q_k A_2(q^2,W), \\
&\textstyle\sum_\beta [(A_j{}^\beta)^* V_k{}^\beta + (V_j{}^\beta)^* A_k{}^\beta] = i\epsilon_{kjl} q_l I(q^2,W), \\
&\textstyle\sum_\beta |V_D{}^\beta|^2 = D_V(q^2,W), \\
&\textstyle\sum_\beta |A_D{}^\beta|^2 = D_A(q^2,W), \\
&\textstyle\sum_\beta [(V_k{}^\beta)^* V_D{}^\beta - (V_D{}^\beta)^* V_k{}^\beta] = iq_k I_V(q^2,W), \\
&\textstyle\sum_\beta [(A_k{}^\beta)^* A_D{}^\beta - (A_D{}^\beta)^* A_k{}^\beta] = iq_k I_A(q^2,W), \\
&\textstyle\sum_\beta [(A_D{}^\beta)^* V_D{}^\beta + (V_D{}^\beta)^* A_D{}^\beta] = 0, \\
&\textstyle\sum_\beta [(V_k{}^\beta)^* A_D{}^\beta - (A_D{}^\beta)^* V_k{}^\beta] = 0, \\
&\textstyle\sum_\beta [(A_k{}^\beta)^* V_D{}^\beta - (V_D{}^\beta)^* A_k{}^\beta] = 0,
\end{aligned} \tag{41}$$

with all the structure functions [V_1, V_2 etc.] in Eq. (41) *real*.

All that remains now is to evaluate the tensor contractions contained in Eq. (37). Using the equations

$$\begin{aligned} q_n(\delta_{nk}-q_nq_k/q_0^2)&=-(q^2/q_0^2)q_k\,,\\ q_nq_mt_{nm}&=2E_\nu E_l(E_\nu-E_l)^2\cos^2(\phi/2)\,,\\ \delta_{nm}t_{nm}&=q^2+2E_\nu E_l\cos^2(\phi/2)\,,\\ \epsilon_{nml}q_lt_{nm}&=iq^2(E_\nu+E_l)\,, \end{aligned} \tag{42}$$

we get, by some straightforward algebra, the result

$$\begin{aligned} d^2\sigma/d\Omega_l dE_l&=[g^2/(2\pi)^2](E_l/E_\nu)\kappa\,,\\ \kappa&=q^2\alpha(q^2,W)+2E_\nu E_l\cos^2(\tfrac{1}{2}\phi)\beta(q^2,W)\\ &\quad-q^2(E_\nu+E_l)\gamma(q^2,W)\,,\\ \alpha(q^2,W)&=V_1(q^2,W)+A_1(q^2,W)\,,\\ \beta(q^2,W)&=\{q^2[V_1(q^2,W)+A_1(q^2,W)]\\ &\quad+(q^2)^2[V_2(q^2,W)+A_2(q^2,W)]\\ &\quad+q^2[I_V(q^2,W)+I_A(q^2,W)]+D_V(q^2,W)\\ &\quad+D_A(q^2,W)\}4M_N^2/(W^2-M_N^2+q^2)^2\,,\\ \gamma(q^2,W)&=I(q^2,W)\,. \end{aligned} \tag{43}$$

The formula for antineutrino-induced reactions is the same, except that the final term in κ is changed to $+q^2(E_\nu+E_l)\gamma(q^2,W)$ [and, of course, in Eq. (32) defining V_k and A_k, the currents J_k^V and J_k^A are replaced by their adjoints].

The simplest illustration of our result is the elastic reaction $\nu+N\to l+N$. Explicit calculation shows that $d^2\sigma(\bar{\nu}+p\to \bar{l}+n)/d\Omega_l dE_l$ has the form of Eq. (13), with

$$\begin{aligned} \alpha^{(-)}(q^2,W)&=\delta(W-M_N)[(1+q^2/4M_N^2)g_A(q^2)^2\\ &\quad+(q^2/4M_N^2)g_V(q^2)^2]\,,\\ \beta^{(-)}(q^2,W)&=\delta(W-M_N)[g_A(q^2)^2\\ &\quad+F_1^V(q^2)^2+q^2F_2^V(q^2)^2]\,,\\ \gamma^{(-)}(q^2,W)&=\delta(W-M_N)[-g_A(q^2)g_V(q^2)/M_N]\,. \end{aligned} \tag{44}$$

We have also computed, for this reaction, the individual structure functions appearing in Eq. (41). They are

$$\begin{aligned} V_1^{(-)}(q^2,W)&=\delta(W-M_N)(q^2/4M_N^2)g_V(q^2)^2\,,\\ V_2^{(-)}(q^2,W)&=\delta(W-M_N)\{[1+q^2/4M_N^2]f_V(q^2)^2\\ &\quad-g_V(q^2)f_V(q^2)/M_N\}\,,\\ A_1^{(-)}(q^2,W)&=\delta(W-M_N)(1+q^2/4M_N^2)g_A(q^2)^2\,,\\ A_2^{(-)}(q^2,W)&=\delta(W-M_N)[(q^2/4M_N^2)h_A(q^2)^2\\ &\quad-h_A(q^2)g_A(q^2)/M_N]\,,\\ I^{(-)}(q^2,W)&=\delta(W-M_N)[-g_A(q^2)g_V(q^2)/M_N]\,,\\ I_A^{(-)}(q^2,W)&=\delta(W-M_N)(-1/2M_N^2)\\ &\quad\times[2M_Ng_A(q^2)-q^2h_A(q^2)]^2\,,\\ D_A^{(-)}(q^2,W)&=\delta(W-M_N)(q^2/4M_N^2)\\ &\quad\times[2M_Ng_A(q^2)-q^2h_A(q^2)]^2\,,\\ I_V^{(-)}(q^2,W)&=D_V^{(-)}(q^2,W)=0\,. \end{aligned} \tag{45}$$

4. DERIVATION OF THE SUM RULES

In this Section we derive the sum rules of Sec. 2. In the first subsection we state and discuss the fundamental identity used in the derivations. In subsequent subsections we derive Eqs. (14), (15), and (16). The derivations for the strangeness-changing case are identical to those for the strangeness-conserving case, and are omitted.

(A) Fundamental Identity

The starting point of the derivations is the identity[6]

$$\begin{aligned} \frac{1}{q_0}\int_0^\infty dt\, e^{iq_0t}\langle N|[A(t),\dot{B}(0)]|N\rangle &=-i\langle N|[A(0),B(0)]|N\rangle\\ &\quad+(2q_0)^{-1}\langle N|[A(0),B(0)]+[\dot{B}(0),A(0)]|N\rangle\\ &\quad+q_0\int_0^\infty dt\, e^{iq_0t}\langle N|[A(t),B(0)]|N\rangle\,, \end{aligned} \tag{46}$$

where

$$\dot{A}(t)=\frac{dA(t)}{dt}\,,\quad \dot{B}(t)=\frac{dB(t)}{dt} \tag{47}$$

are the time derivatives of $A(t)$ and $B(t)$. Equation (46) is easily derived by repeated integration by parts, and holds for all q_0 in the upper half of the complex plane. In this paper, the operators $A(t)$ and $B(t)$ will always be of the form

$$\begin{aligned} A(t)&=-i\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\mathfrak{F}_A(\mathbf{x},t)\,,\\ B(t)&=-i\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\mathfrak{F}_B(\mathbf{y},t)\,;\\ \mathfrak{F}_A&=\mathfrak{F}_{a\lambda}\text{ or }\mathfrak{F}_{a\lambda}{}^5\,,\quad \mathfrak{F}_B=\mathfrak{F}_{b\sigma}\text{ or }\mathfrak{F}_{b\sigma}{}^5\,. \end{aligned} \tag{48}$$

In (I) we studied Eq. (46) with $\mathbf{s}=0$; this led, in the limit $q_0\to 0$, to sum rules at $q^2=0$. In this paper we will study the case when $\mathbf{s}\neq 0$, and will find, in the limit as $q_0\to 0$, sum rules for fixed q^2 (with $q^2=|\mathbf{s}|^2$).

There are a number of features which all of the derivations given below have in common. First of all, we will always use Eq. (46) with the nucleon N at rest, and with the nucleon spin averaged over. Secondly, each term of Eq. (46) can be divided into a part which is symmetric and a part which is antisymmetric in the unitary spin indices a and b. We will only study the identity for the *antisymmetric parts*. In each case below, we will find that the term

$$U=(2q_0)^{-1}\langle N|[\dot{A}(0),B(0)]+[\dot{B}(0),A(0)]|N\rangle \tag{49}$$

[6] Equation (46) is a more symmetrical version of Eq. (37) of Ref. 2. Equation (46) remains valid if $\langle N|$ and $|N\rangle$ are replaced by any two states of equal four-momentum.

is purely symmetric in the unitary spin indices, and thus makes no contribution. Thirdly, since we have

$$q_0\int_0^\infty dt\, e^{iq_0t}\langle N|[A(t),B(0)]|N\rangle$$
$$=-iq_0\sum_{\substack{\beta,\mathrm{INT}\\(M_\beta\geq M_N)}}\left\{\frac{\langle N|\mathfrak{F}_A|\beta\rangle\langle\beta|\mathfrak{F}_B|N\rangle}{q_0+M_N-(|\mathbf{s}|^2+M_\beta{}^2)^{1/2}}-\frac{\langle N|\mathfrak{F}_B|\beta\rangle\langle\beta|\mathfrak{F}_A|N\rangle}{q_0+(|\mathbf{s}|^2+M_\beta{}^2)^{1/2}-M_N}\right\}(2\pi)^3\delta(0),\quad(50)$$

the limit as $q_0\to0$ of Eq. (50) is zero for all $|\mathbf{s}|^2>0$. As a result, the third term on the right-hand side of Eq. (46) makes no contribution to the sum rules.[7]

Finally, we will always find that the unitary spin-antisymmetric part of

$$\int_0^\infty dt\, e^{iq_0t}\langle N|[A(t),\dot{B}(0)]|N\rangle\quad(51)$$

is an odd function of q_0, $O(q_0,q^2)$. Thus, in the limit as $q_0\to0$ the identity of Eq. (46) will become the equation

$$\frac{\partial}{\partial q_0}O(q_0,q^2)\bigg|_{q_0=0}=C,\quad(52)$$

where C is the unitary spin-antisymmetric part of the commutator $-i\langle N|[A(0),B(0)]|N\rangle$. Equation (52) states that the commutator of A and B is related to the energy derivative of a forward scattering amplitude, evaluated at zero energy. Up to this point the derivation is *rigorous*. Now, in order to relate the left-hand side of Eq. (52) to physically measurable quantities, we will assume that the energy derivative $(\partial/\partial q_0)O(q_0,q^2)$ satisfies an *unsubtracted* dispersion relation in the energy variable q_0, for fixed q^2. The discontinuity of $(\partial/\partial q_0)\times O(q_0,q^2)$ across its cuts will, in each case considered, be related to the structure functions defined in Eq. (41).

(B) Sum Rule for $\beta^{(\pm)}$

The sum rule on $\beta^{(\pm)}$ of Eq. (14) is obtained by adding together two separately derived sum rules on the axial-vector and the vector parts of $\beta^{(\pm)}$, $\beta_A{}^{(\pm)}$, and $\beta_V{}^{(\pm)}$:

$$1=g_A(q^2)^2+\int_{M_N+M_\pi}^\infty\frac{W}{M_N}dW\times[\beta_A{}^{(-)}(q^2,W)-\beta_A{}^{(+)}(q^2,W)],\quad(53\text{a})$$

$$1=F_1{}^V(q^2)^2+q^2F_2{}^V(q^2)^2+\int_{M_N+M_\pi}^\infty\frac{W}{M_N}dW\times[\beta_V{}^{(-)}(q^2,W)-\beta_V{}^{(+)}(q^2,W)].\quad(53\text{b})$$

In terms of the structure functions defined in Eq. (41),

$$\beta_A{}^{(\pm)}(q^2,W)=[q^2A_1{}^{(\pm)}(q^2,W)+(q^2)^2A_2{}^{(\pm)}(q^2,W)+q^2I_A{}^{(\pm)}(q^2,W)+D_A{}^{(\pm)}(q^2,W)]\times4M_N{}^2/(W^2-M_N{}^2+q^2)^2,\quad(54)$$
$$\beta_V{}^{(\pm)}(q^2,W)=q^2[V_1{}^{(\pm)}(q^2,W)+q^2V_2{}^{(\pm)}(q^2,W)]\times4M_N{}^2/(W^2-M_N{}^2+q^2)^2.$$

[The structure functions $I_V{}^{(\pm)}(q^2,W)$ and $D_V{}^{(\pm)}(q^2,W)$ vanish identically in the strangeness-conserving case, because of conservation of the vector current.] Since the derivations of Eqs. (53a) and (53b) are identical, we will treat explicitly only the axial-vector case, Eq. (53a).

We start from the fundamental identity of Eq. (46), taking

$$A(t)=-i\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\mathfrak{F}_{a4}{}^5(\mathbf{x},t),$$
$$B(t)=-i\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\mathfrak{F}_{b4}{}^5(\mathbf{y},t).\quad(55)$$

Defining $D_a(x)=\partial_\lambda\mathfrak{F}_{a\lambda}{}^5(x)$ we find, by spatial integration by parts, that

$$\dot{A}(t)=\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}[D_a(\mathbf{x},t)-is_n\mathfrak{F}_{an}{}^5(\mathbf{x},t)],$$
$$\dot{B}(t)=\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}[D_b(\mathbf{y},t)+is_n\mathfrak{F}_{bn}{}^5(\mathbf{y},t)],\quad(56)$$

where the repeated index n is summed over. With A and B as shown in Eq. (55), the first term on the right-hand side of Eq. (46) becomes, using the local commutation relation of Eq. (1c),

$$-i\sum_s\langle N|[A(0),B(0)]|N\rangle=\epsilon_{abc}\langle\tfrac{1}{2}\tau^c\rangle(2\pi)^3\delta(0).\quad(57)$$

Thus this term is purely antisymmetric in the isospin indices a and b. [Note that the validity of Eq. (57) depends on the correctness of the local commutation relation. If Eq. (1c) were modified by the addition of a term proportional to $\nabla^2\delta(\mathbf{x}-\mathbf{y})$, a term proportional to $|\mathbf{s}|^2$ would be added to Eq. (57).] The second term on the right-hand side of Eq. (46) becomes

$$U_1{}^{ab}=\frac{-1}{2q_0}\left\{\sum_s\langle N|\left[\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\frac{\partial\mathfrak{F}_{a4}{}^5(\mathbf{x},t)}{\partial t},\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\mathfrak{F}_{b4}{}^5(\mathbf{y},t)\right]|N\rangle+\sum_s\langle N|\left[\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\frac{\partial\mathfrak{F}_{b4}{}^5(\mathbf{y},t)}{\partial t},\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\mathfrak{F}_{a4}{}^5(\mathbf{x},t)\right]|N\rangle\right\}\quad(58\text{a})$$

[7] Only when $|\mathbf{s}|^2=0$ does the one-nucleon intermediate state ($M_\beta=M_N$) make a contribution to the limit. This is the case considered in Ref. 2.

$$=\frac{-1}{2q_0}\left\{\sum_s\langle N|\left[\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\frac{\partial\mathfrak{F}_{a4}{}^5(\mathbf{x},t)}{\partial t},\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\mathfrak{F}_{b4}{}^5(\mathbf{y},t)\right]|N\rangle\right.$$
$$\left.+\sum_s\langle N|\left[\int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}}\frac{\partial\mathfrak{F}_{b4}{}^5(\mathbf{x},t)}{\partial t},\int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}}\mathfrak{F}_{a4}{}^5(\mathbf{y},t)\right]|N\rangle\right\},\qquad(58b)$$

where we have obtained Eq. (58b) by setting $-\mathbf{y}\leftrightarrow\mathbf{x}$ in the second term of Eq. (58a) and by using the parity transformation properties of the axial-vector current. Clearly $U_1{}^{ab}$ is explicitly symmetric in a and b. Thus, if we agree to keep only the part of Eq. (46) which is antisymmetric in a and b, the second term on the right-hand side of Eq. (46), which involves the unknown commutator of $\partial\mathfrak{F}_{a4}{}^5/\partial t$ with $\mathfrak{F}_{b4}{}^5$, drops out. As discussed above, the limit as $q_0\rightarrow 0$ of the third term on the right-hand side of Eq. (46) vanishes.

Now let us turn to the left-hand side of Eq. (46). Using translational invariance, the integral over $\mathbf{y}$ can be done explicitly, giving an over-all factor of $(2\pi)^3\delta(\mathbf{0})$. We cancel this against the identical factor in Eq. (57). Taking N to be a proton at rest, and multiplying Eq. (46) by an over-all factor ϵ_{ab3} gives

$$1=\frac{\epsilon_{ab3}}{q_0}\int d^4x\exp(-i\hat{q}\cdot x)\theta(x_0)\sum_s\langle p|[D_a(x)-is_n\mathfrak{F}_{an}{}^5(x),D_b(0)+is_m\mathfrak{F}_{bm}{}^5(0)]|p\rangle+o(q_0),\qquad(59a)$$

$$\hat{q}=(\mathbf{s},iq_0),\qquad(59b)$$

where $o(q_0)$ indicates terms which vanish as $q_0\rightarrow 0$. Let us define the amplitudes $d(q_0,q^2)$, $a_1(q_0,q^2)$, $a_2(q_0,q^2)$, and $i_A(q_0,q^2)$ by the equations

$$\begin{aligned}d(q_0,q^2)&=\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\theta(x_0)\sum_s\langle p|[D_a(x),D_b(0)]|p\rangle,\\ a_1(q_0,q^2)\delta_{nm}+a_2(q_0,q^2)q_nq_m&=\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\theta(x_0)\sum_s\langle p|[\mathfrak{F}_{an}{}^5(x),\mathfrak{F}_{bm}{}^5(0)]|p\rangle,\\ iq_ni_A(q_0,q^2)&=\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\theta(x_0)\sum_s\langle p|[\mathfrak{F}_{an}{}^5(x),D_b(0)]-[D_a(x),\mathfrak{F}_{bn}{}^5(0)]|p\rangle.\end{aligned}\qquad(60)$$

We will prove below that these are all odd functions of q_0. Thus Eq. (59), in the limit as $q_0\rightarrow 0$, becomes the statement

$$\begin{aligned}1&=\frac{\partial}{\partial q_0}\lambda(q_0,q^2)\Big|_{q_0=0},\\ \lambda(q_0,q^2)&=d(q_0,q^2)+q^2a_1(q_0,q^2)+(q^2)^2a_2(q_0,q^2)+q^2i_A(q_0,q^2),\end{aligned}\qquad(61)$$

with q^2 fixed at $|\mathbf{s}|^2$.

Let us now study the properties of the functions d, a_1, a_2, and i_A. From their definitions as retarded commutators, it follows by the standard methods of forward dispersion relations[8] that they are analytic functions of q_0 in the upper half q_0 plane, for fixed q^2. Thus if we assume that the amplitude $(\partial/\partial q_0)\lambda(q_0,q^2)$ approaches zero as $q_0\rightarrow\infty$ in the upper half plane, we can write the unsubtracted dispersion relation

$$\frac{\partial}{\partial q_0}\lambda(q_0,q^2)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dq_0'}{(q_0'-q_0)^2}\{d'(q_0',q^2)+q^2a_1'(q_0',q^2)+(q^2)^2a_2'(q_0',q^2)+q^2i_A'(q_0',q^2)\},\qquad(62)$$

where the absorptive parts d', a_1', a_2', i_A' are defined by

$$\begin{aligned}id'(q_0,q^2)&=\tfrac{1}{2}\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\sum_s\langle p|[D_a(x),D_b(0)]|p\rangle,\\ i[a_1'(q_0,q^2)\delta_{nm}+a_2'(q_0,q^2)q_nq_m]&=\tfrac{1}{2}\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\sum_s\langle p|[\mathfrak{F}_{an}{}^5(x),\mathfrak{F}_{bm}{}^5(0)]|p\rangle,\\ i[iq_ni_A'(q_0,q^2)]&=\tfrac{1}{2}\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\sum_s\langle p|[\mathfrak{F}_{an}{}^5(x),D_b(0)]-[D_a(x),\mathfrak{F}_{bn}{}^5(0)]|p\rangle.\end{aligned}\qquad(63)$$

[8] J. D. Jackson, *Dispersion Relations*, edited by G. R. Screaton (Interscience Publishers, Inc., New York, 1961), pp. 1–32.

The next step is to evaluate the absorptive parts. Let us consider explicitly the case of d'. Let $k_p=(0,iM_N)$ be the proton four-momentum. Inserting a complete set of intermediate states, we find that

$$id'(q_0,q^2)=\tfrac{1}{2}\epsilon_{ab3}(2\pi)^4\sum_{\beta,\mathrm{INT}}\sum_s\int\frac{d^3k_\beta}{(2\pi)^3}$$
$$\times[\langle p|D_a(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_b(0)|p\rangle\delta(k_\beta-q-k_p)-\langle p|D_b(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_a(0)|p\rangle\delta(k_\beta+q-k_p)]$$
$$=\pi\epsilon_{ab3}\sum_{\beta,\mathrm{INT}}\sum_s\{[\langle p|D_a(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_b(0)|p\rangle]|_{\mathbf{k}_\beta=\mathbf{q}}\delta(k_{\beta0}-q_0-M_N)$$
$$-[\langle p|D_b(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_a(0)|p\rangle]|_{\mathbf{k}_\beta=-\mathbf{q}}\delta(k_{\beta0}+q_0-M_N)\}. \quad (64)$$

Parity invariance tells us that

$$\sum_{\beta,\mathrm{INT}}\sum_s[\langle p|D_b(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_a(0)|p\rangle]|_{\mathbf{k}_\beta=-\mathbf{q}}\delta(k_{\beta0}+q_0-M_N)$$
$$=\sum_{\beta,\mathrm{INT}}\sum_s[\langle p|D_b(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_a(0)|p\rangle]|_{\mathbf{k}_\beta=\mathbf{q}}\delta(k_{\beta0}+q_0-M_N). \quad (65)$$

Thus Eq. (64) can be written, using the antisymmetry of ϵ_{ab3}, as

$$id'(q_0,q^2)=\pi\epsilon_{ab3}\sum_{\beta,\mathrm{INT}}\sum_s[\langle p|D_a(0)|\beta(k_\beta)\rangle\langle\beta(k_\beta)|D_b(0)|p\rangle]|_{\mathbf{k}_\beta=\mathbf{q}}[\delta(k_{\beta0}-q_0-M_N)+\delta(k_{\beta0}+q_0-M_N)]. \quad (66)$$

We see that d' is an even function of q_0; hence d is an odd function of q_0. Since

$$\epsilon_{ab3}D_a{}^*D_b=D_1{}^*D_2-D_2{}^*D_1=\tfrac{1}{2}i[(D_1{}^*+iD_2{}^*)(D_1-iD_2)-(D_1{}^*-iD_2{}^*)(D_1+iD_2)], \quad (67)$$

we obtain finally the result that

$$d'(q_0,q^2)=\tfrac{1}{2}\pi[D^{(-)}-D^{(+)}],\quad q_0>0, \quad (68)$$

with

$$D^{(-)}=\sum_{\beta,\mathrm{INT}}\sum_s|\langle\beta[\mathbf{q},i(q_0+M_N)]|D_1(0)-iD_2(0)|p\rangle|^2\delta(k_{\beta0}-q_0-M_N),$$
$$D^{(+)}=\sum_{\beta,\mathrm{INT}}\sum_s|\langle\beta[\mathbf{q},i(q_0+M_N)]|D_1(0)+iD_2(0)|p\rangle|^2\delta(k_{\beta0}-q_0-M_N). \quad (69)$$

Clearly Eq. (69) is identical with Eqs. (41), (32), and (33), defining the structure function D, with q_0 given by

$$q_0=E_\nu-E_l=(W^2-M_N{}^2+q^2)/2M_N. \quad (70)$$

In a similar manner we find that a_1', a_2', and i_A' are even functions of q_0 (which implies that a_1, a_2, and i_A are odd functions of q_0). Also, we find that for $q_0>0$,

$$a_1'(q_0,q^2)=\tfrac{1}{2}\pi[A_1{}^{(-)}-A_1{}^{(+)}],\quad a_2'(q_0,q^2)=\tfrac{1}{2}\pi[A_2{}^{(-)}-A_2{}^{(+)}],\quad i_A'(q_0,q^2)=\tfrac{1}{2}\pi[I_A{}^{(-)}-I_A{}^{(+)}], \quad (71)$$

where the structure functions $A_1{}^{(\pm)}$, $A_2{}^{(\pm)}$, and $I_A{}^{(\pm)}$ are those defined in Eq. (41). Combining Eqs. (43), (61), (62), and (68)–(71), we see that we have derived the sum rule

$$1=\int dq_0[\beta_A{}^{(-)}-\beta_A{}^{(+)}]=\int\frac{W}{M_N}dW[\beta_A{}^{(-)}(q^2,W)-\beta_A{}^{(+)}(q^2,W)]. \quad (72)$$

Using Eq. (44), the pole contribution to Eq. (72) can be explicitly evaluated, giving Eq. (53a).

(C) Sum Rule for $\alpha^{(\pm)}$

The sum rule on $\alpha^{(\pm)}$ of Eq. (15) is obtained by adding together the two identities

$$C_I{}^2=\left(1+\frac{q^2}{4M_N{}^2}\right)g_A(q^2)^2+\int_{M_N+M_\pi}^\infty\frac{W}{M_N}dW[\alpha_A{}^{(-)}(q^2,W)-\alpha_A{}^{(+)}(q^2,W)], \quad (73a)$$

$$C_I{}^1=\left(\frac{q^2}{4M_N{}^2}\right)g_V(q^2)^2+\int_{M_N+M_\pi}^\infty\frac{W}{M_N}dW[\alpha_V{}^{(-)}(q^2,W)-\alpha_V{}^{(+)}(q^2,W)]. \quad (73b)$$

Here $\alpha_A{}^{(\pm)}$ and $\alpha_V{}^{(\pm)}$ are, respectively, the axial-vector and the vector parts of $\alpha^{(\pm)}$,

$$\alpha_A{}^{(\pm)}=A_1{}^{(\pm)}(q^2,W),\quad \alpha_V{}^{(\pm)}=V_1{}^{(\pm)}(q^2,W). \quad (74)$$

We will sketch the derivation of Eq. (73a); the derivation of Eq. (73b) is identical.

In order to derive Eq. (73a), we use the fundamental identity, with

$$A(t)=-i\int d^3x\, e^{-is\cdot x}\mathfrak{F}_{an}{}^5(\mathbf{x},t)\,,\quad B(t)=-i\int d^3y\, e^{is\cdot y}\mathfrak{F}_{bm}{}^5(\mathbf{y},t)\,. \tag{75}$$

Using Eqs. (4a) and (11), the first term on the right-hand side of Eq. (46) becomes

$$-i\sum_s\langle p|[A(0),B(0)]|p\rangle=-\epsilon_{abc}\delta_{nm}C_I{}^2\langle\tfrac{1}{2}\tau^c\rangle(2\pi)^3\delta(\mathbf{0})+(\text{symmetric in } ab)\,. \tag{76}$$

The second term is

$$U_2{}^{nm,ab}=\frac{-1}{2q_0}\sum_s\langle p|\int d^3x\, e^{-is\cdot x}\int d^3y\, e^{is\cdot y}\left\{\left[\frac{\partial\mathfrak{F}_{an}{}^5(\mathbf{x},t)}{\partial t},\,\mathfrak{F}_{bm}{}^5(\mathbf{y},t)\right]+\left[\frac{\partial\mathfrak{F}_{bm}{}^5(\mathbf{y},t)}{\partial t},\,\mathfrak{F}_{an}{}^5(\mathbf{x},t)\right]\right\}|p\rangle, \tag{77}$$

which, by using the parity transformation properties of $\mathfrak{F}^5$, is equal to

$$\frac{-1}{2q_0}\sum_s\langle p|\int d^3x\, e^{-is\cdot x}\int d^3y\, e^{is\cdot y}\left\{\left[\frac{\partial\mathfrak{F}_{an}{}^5(\mathbf{x},t)}{\partial t},\,\mathfrak{F}_{bm}{}^5(\mathbf{y},t)\right]+\left[\frac{\partial\mathfrak{F}_{bm}{}^5(\mathbf{x},t)}{\partial t},\,\mathfrak{F}_{an}{}^5(\mathbf{y},t)\right]\right\}|p\rangle. \tag{78}$$

The expression in Eq. (78) is explicitly symmetric under the simultaneous interchanges $n\leftrightarrow m$, $a\leftrightarrow b$. Since parity invariance requires that U_2 be of the form

$$U_2{}^{nm,ab}=\mu_1{}^{ab}\delta_{nm}+\mu_2{}^{ab}s_ns_m\,, \tag{79}$$

U_2 is symmetric under the interchange $a\leftrightarrow b$. Thus, if we keep only terms which are antisymmetric in a and b, the unwanted $[\partial\mathfrak{F}^5/\partial t,\mathfrak{F}^5]$ terms drop out.

As a result, we are left with the identity

$$\begin{gathered}\delta_{nm}C_I{}^2=\frac{\partial}{\partial q_0}\eta(q_0,q^2)\Big|_{q_0=0}\,,\\ \eta(q_0,q^2)=\bar{a}_1(q_0,q^2)\delta_{nm}+\bar{a}_2(q_0,q^2)q_nq_m=\epsilon_{ab3}\int d^4x\, e^{-iq\cdot x}\theta(x_0)\sum_s\langle p|\left[\frac{\partial\mathfrak{F}_{an}{}^5(x)}{\partial x_0},\frac{\partial\mathfrak{F}_{bm}{}^5(0)}{\partial t}\right]|p\rangle.\end{gathered} \tag{80}$$

[Here $\partial\mathfrak{F}_{bm}{}^5(0)/\partial t$ denotes $\partial\mathfrak{F}_{bm}{}^5(\mathbf{y},t)/\partial t$ evaluated at $\mathbf{y}=\mathbf{0}$, $t=0$.] Let us now postulate that

$$\partial\bar{a}_1(q_0,q^2)/\partial q_0 \tag{81}$$

satisfies an unsubtracted dispersion relation. It is easy to see that the absorptive part of $\bar{a}_1(q_0,q^2)$ is just $q_0{}^2$ times the absorptive part of the amplitude $a_1(q_0,q^2)$ defined in Eq. (60). Thus, the δ_{nm} term in Eq. (80) becomes

$$C_I{}^2=\int dq_0(A_1{}^{(-)}-A_1{}^{(+)})=\int\frac{W}{M_N}dW[A_1{}^{(-)}(q^2,W)-A_1{}^{(+)}(q^2,W)], \tag{82}$$

which is the result to be proved.

(D) Sum Rule for $\gamma^{(\pm)}$

The sum rule on $\gamma^{(\pm)}$ of Eq. (16) is derived by adding the fundamental identity, with

$$A_1(t)=-i\int d^3x\, e^{-is\cdot x}\mathfrak{F}_{an}{}^5(\mathbf{x},t)\,,\quad B_1(t)=-i\int d^3y\, e^{is\cdot y}\mathfrak{F}_{bm}(\mathbf{y},t)\,, \tag{83}$$

to the same identity, with

$$A_2(t)=-i\int d^3x\, e^{-is\cdot x}\mathfrak{F}_{an}(\mathbf{x},t)\,,\quad B_2(t)=-i\int d^3y\, e^{is\cdot y}\mathfrak{F}_{bm}{}^5(\mathbf{y},t)\,. \tag{84}$$

Using Eq. (4b), the first term on the right-hand side of Eq. (46) is

$$-i\sum_s\langle p|[A_1(0),B_1(0)]+[A_2(0),B_2(0)]|p\rangle=(\text{symmetric in } ab)\,, \tag{85}$$

since

$$\sum_s \langle p | \mathcal{A}_{c4}(0) | p \rangle = 0 \tag{86}$$

for nucleon states at rest. The second term, using the parity transformation properties of the currents, becomes

$$U_3^{nm,ab} = \frac{-1}{2q_0} \sum_s \langle p | \int d^3x\, e^{-i\mathbf{s}\cdot\mathbf{x}} \int d^3y\, e^{i\mathbf{s}\cdot\mathbf{y}} \Bigg\{ \left[\frac{\partial \mathfrak{F}_{an}^5(\mathbf{x},t)}{\partial t}, \mathfrak{F}_{bm}(\mathbf{y},t) \right] - \left[\frac{\partial \mathfrak{F}_{bm}(\mathbf{x},t)}{\partial t}, \mathfrak{F}_{an}^5(\mathbf{y},t) \right]$$
$$+ \left[\frac{\partial \mathfrak{F}_{an}(\mathbf{x},t)}{\partial t}, \mathfrak{F}_{bm}^5(\mathbf{y},t) \right] - \left[\frac{\partial \mathfrak{F}_{bm}^5(\mathbf{x},t)}{\partial t}, \mathfrak{F}_{an}(\mathbf{y},t) \right] \Bigg\} | p \rangle$$
$$= \mu_3^{ab} \epsilon_{nml} s_l . \tag{87}$$

Clearly, μ_3^{ab} is symmetric in a and b. If we keep only the antisymmetric part of the identity, the $[\partial \mathfrak{F}^5/\partial t, \mathfrak{F}]$ and the $[\partial \mathfrak{F}/\partial t, \mathfrak{F}^5]$ terms drop out.

Thus, we get the identity

$$0 = \frac{\partial}{\partial q_0} i(q_0, q^2) \Bigg|_{q_0=0} ,$$
$$i(q_0,q^2) = \epsilon_{ab3} \int d^4x\, e^{-iq\cdot x} \theta(x_0) \sum_s \langle p | \left[\frac{\partial \mathfrak{F}_{an}^5(x)}{\partial x_0}, \frac{\partial \mathfrak{F}_{bm}(0)}{\partial t} \right] + \left[\frac{\partial \mathfrak{F}_{an}(x)}{\partial x_0}, \frac{\partial \mathfrak{F}_{bm}^5(0)}{\partial t} \right] | p \rangle . \tag{88}$$

The postulate that

$$\partial i(q_0,q^2)/\partial q_0 \tag{89}$$

satisfies an unsubtracted dispersion relation in q_0 leads immediately to Eq. (16).

ACKNOWLEDGMENTS

I am grateful to Professor M. Gell-Mann for a discussion essential to the genesis of this paper, and to Dr. C. G. Callan for an interesting conversation. I wish to thank Professor L. Van Hove for the hospitality of the CERN Theoretical Study Division during the summer of 1965.

APPENDIX

In this Appendix we give the generalization of the results stated in Sec. 2 to the case when all lepton-mass terms are included. In order to calculate lepton-mass corrections, it is easier to work covariantly, rather than to eliminate the fourth components of currents in terms of spatial components and divergences. Thus we write

$$T_{\lambda\sigma} = \sum_{\beta,\mathrm{INT}} \sum_s \delta(k_{\beta 0} - k_{N0} - q_0) \langle N(k_N) | (J_\sigma^V + J_\sigma^A)^* | \beta(k_N+q) \rangle \langle \beta(k_N+q) | J_\lambda^V + J_\lambda^A | N(k_N) \rangle$$
$$= \frac{M_N}{k_{N0}} [\bar{A}\delta_{\lambda\sigma} + \bar{B} k_{N\lambda} k_{N\sigma} + \bar{C} \epsilon_{\lambda\sigma\gamma\delta} q_\gamma k_{N\delta} + \bar{D} q_\lambda q_\sigma + \bar{E}(q_\lambda k_{N\sigma} + q_\sigma k_{N\lambda})] , \tag{A1}$$

with $\bar{A}, \cdots, \bar{E}$ functions of q^2 and W. Time reversal and parity invariance rule out the presence of a term proportional to $q_\lambda k_{N\sigma} - q_\sigma k_{N\lambda}$ in Eq. (A1). Comparing Eq. (A1) with Eq. (41), in the laboratory frame, shows that

$$\begin{aligned} &\bar{A} = \alpha(q^2,W) , \quad M_N^2 \bar{B} = \beta(q^2,W) , \quad M_N \bar{C} = \gamma(q^2,W) , \quad \bar{D} \equiv \delta(q^2,W) = V_2(q^2,W) + A_2(q^2,W) , \\ &M_N \bar{E} \equiv \epsilon(q^2,W) = q_0^{-1} \{ V_1(q^2,W) + A_1(q^2,W) + q^2 [V_2(q^2,W) + A_2(q^2,W)] + \tfrac{1}{2} [I_V(q^2,W) + I_A(q^2,W)] \} . \end{aligned} \tag{A2}$$

It is straightforward to calculate the contraction of $T_{\lambda\sigma}$ with the leptonic trace. We find that Eq. (13) and Eq. (22) for the strangeness-conserving case are replaced by

$$d^2\sigma\left(\binom{\nu}{\bar{\nu}} + p \to \binom{l}{\bar{l}} + \beta(S=0) \right) \Big/ d\Omega_l dE_l = \frac{G^2 \cos^2\theta_C}{(2\pi)^2} \frac{[(E_\nu - q_0)^2 - m_l^2]^{1/2}}{E_\nu} \kappa^{(\pm)} , \tag{A3}$$

$$d^2\sigma\left(\binom{\nu}{\bar{\nu}} + p \to \binom{l}{\bar{l}} + \beta(S=0) \right) \Big/ d(q^2) dq_0 = \frac{G^2 \cos^2\theta_C}{4\pi E_\nu^2} \kappa^{(\pm)} , \tag{A4}$$

with

$$\kappa^{(\pm)}=(q^2+m_l^2)\alpha^{(\pm)}(q^2,W)+[2E_\nu^2-2E_\nu q_0-\tfrac{1}{2}(q^2+m_l^2)]\beta^{(\pm)}(q^2,W)$$
$$\mp[(2E_\nu-q_0)q^2-m_l^2q_0]\gamma^{(\pm)}(q^2,W)+\tfrac{1}{2}m_l^2(q^2+m_l^2)\delta^{(\pm)}(q^2,W)-2m_l^2E_\nu\epsilon^{(\pm)}(q^2,W). \quad \text{(A5)}$$

Inspection of Eq. (A5) and its analog for a neutron target shows that $\beta^{(\pm)}$, $\gamma^{(\pm)}$, $\epsilon^{(\pm)}$, and $\alpha^{(\pm)}+\frac{1}{2}m_l^2\delta^{(\pm)}$ are independently measurable. Since the derivation of the sum rule on $\alpha^{(\pm)}$ given in Sec. 4 shows that

$$0=\int dq_0(\delta^{(-)}-\delta^{(+)}), \quad \text{(A6)}$$

we may modify Eq. (15) to read

$$C_I^1+C_I^2=(1+q^2/4M_N^2)g_A(q^2)^2+(q^2/4M_N^2)g_V(q^2)^2$$
$$+\tfrac{1}{2}m_l^2[(1+q^2/4M_N^2)f_V(q^2)^2-g_V(q^2)f_V(q^2)/M_N+(q^2/4M_N^2)h_A(q^2)^2-h_A(q^2)g_A(q^2)/M_N]$$
$$+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}dW[\alpha^{(-)}(q^2,W)+\tfrac{1}{2}m_l^2\delta^{(-)}(q^2,W)-\alpha^{(+)}(q^2,W)-\tfrac{1}{2}m_l^2\delta^{(+)}(q^2,W)]. \quad \text{(A7)}$$

Thus, in the strangeness-conserving case, when lepton-mass terms are included there are still three sum rules which may be directly compared with experiment.

In the strangeness-changing case, equations similar to Eqs. (A3)–(A5) hold, and $\beta_{(p,n)}{}^{(\pm)}$, $m_l^2\epsilon_{(p,n)}{}^{(\pm)}\pm q^2\gamma_{(p,n)}{}^{(\pm)}$, and $\alpha_{(p,n)}{}^{(\pm)}+\frac{1}{2}m_l^2\delta_{(p,n)}{}^{(\pm)}\pm q_0\gamma_{(p,n)}{}^{(\pm)}$ are independently measurable. We see that in this case, when lepton-mass terms are included, only the sum rules on $\beta_{(p,n)}{}^{(\pm)}$ can be directly compared with experiment.

It is easy to verify that the results of Eq. (26) and Eq. (27), referring to the high neutrino-energy behavior of neutrino cross sections, are unchanged by adding the lepton-mass terms. Equation (24) becomes

$$\frac{d\sigma}{dq^2}=\int_{(q^2/2M_N)-}^{E_\nu(1-L/4E_\nu^2)}dq_0\frac{d^2\sigma}{d(q^2)dq_0}, \quad L=q^2+m_l^2+\frac{4E_\nu^2m_l^2}{q^2+m_l^2}. \quad \text{(A8)}$$

If, in addition to Eq. (25), we postulate that

$$\int^\infty\frac{dq_0}{q_0^2}(\delta^{(-)}-\delta^{(+)}), \quad \int^\infty\frac{dq_0}{q_0}(\epsilon^{(-)}-\epsilon^{(+)}) \quad \text{(A9)}$$

are convergent (and similarly in the strangeness-changing case), then we immediately obtain Eqs. (26) and (27).

Sum Rules Giving Tests of Local Current Commutation Relations in High-Energy Neutrino Reactions, STEPHEN L. ADLER [Phys. Rev. 143, 1144 (1966)]. In Eq. (45) for $V_2^{(-)}(q^2,W)$, in both terms on the right-hand side, $f_V(q^2)$ should be $F_2^V(q^2)$.

PHYSICAL REVIEW VOLUME 156, NUMBER 5 25 APRIL 1967

Neutrino or Electron Energy Needed for Testing Current Commutation Relations*

STEPHEN L. ADLER†
Institute for Advanced Study, Princeton, New Jersey

AND

FREDERICK J. GILMAN‡
California Institute of Technology, Pasadena, California

(Received 19 October 1966)

For small leptonic invariant 4-momentum transfer q^2, we investigate what minimum neutrino or electron energy is needed to test current commutation relations. We find that at laboratory energies of order 5 BeV, it is reasonable to start trying to check the equalities or inequalities implied by the current algebra.

RECENTLY Gell-Mann[1] has postulated that the fourth components of the vector and axial-vector weak-current octets satisfy the local equal-time commutation relations

$$[\mathfrak{F}_{a4}(x),\mathfrak{F}_{b4}(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}(x)\delta(\mathbf{x}-\mathbf{y}), \quad (1a)$$

$$[\mathfrak{F}_{a4}(x),\mathfrak{F}_{b4}{}^5(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}{}^5(x)\delta(\mathbf{x}-\mathbf{y}), \quad (1b)$$

$$[\mathfrak{F}_{a4}{}^5(x),\mathfrak{F}_{b4}{}^5(y)]|_{x_0=y_0}=-f_{abc}\mathfrak{F}_{c4}(x)\delta(\mathbf{x}-\mathbf{y}). \quad (1c)$$

It has been shown by Adler[2] that Eqs. (1a) and (1c) can be tested in high-energy neutrino reactions, where they imply that, in the limit of infinite incident neutrino energy, the difference $d\sigma_T(\bar{\nu}+N)/dq^2-d\sigma_T(\nu+N)/dq^2$ approaches a constant which is independent of the leptonic invariant 4-momentum transfer q^2. Bjorken,[3] by an isospin rotation, has transformed the neutrino-reaction results into inequalities on electron-nucleon (or muon-nucleon) scattering which hold in the limit of infinite incident electron energy; these inequalities may make it feasible to test Eq. (1a) in the near future.

However, before proceeding to experiments, one must answer the question, what is *effectively* an infinite incident neutrino or electron energy E? More precisely, what is the energy $E(q^2,\delta)$ such that for $E>E(q^2,\delta)$ the neutrino equalities hold to within fractional error δ? No general answer to this question can be given. However, we will show, in this paper, that the experimental information used in evaluating the sum rules for the axial-vector coupling constant and for the nucleon isovector radius and magnetic moment suffice to determine $E(q^2=0,\delta)$ in two cases.

The results indicate that at incident energies of order 5 BeV it is reasonable to start trying to check the equalities or inequalities implied by the current algebra, at least for small q^2. It is encouraging that this energy range is accessible to the Stanford Linear Accelerator.

Let us review the results[2] for the high-energy neutrino reaction

$$\nu+N\rightarrow l+\beta. \quad (2)$$

We neglect the lepton mass m_l, and define the following kinematic quantities (noncovariant quantities always refer to the *laboratory frame*):

$$\begin{aligned}
M_N&=\text{nucleon mass},\\
E&=\text{incident neutrino energy},\\
E'&=\text{final lepton energy},\\
k&=\text{neutrino 4-momentum},\\
k'&=\text{final lepton 4-momentum},\\
q^2&=(k-k')^2=\text{leptonic invariant 4-momentum transfer},\\
q_0&=E-E'=\text{leptonic energy transfer},\\
W&=\text{invariant mass of system }\beta\\
&=(2M_Nq_0+M_N{}^2-q^2)^{1/2},\\
q_0&=(W^2-M_N{}^2+q^2)/(2M_N).
\end{aligned} \quad (3)$$

Let G be the Fermi coupling constant and θ_c the Cabibbo angle. Then the cross section for strangeness-zero ($S=0$) final states may be written

$$d^2\sigma\left[\binom{\nu}{\bar{\nu}}+p\rightarrow\binom{l}{\bar{l}}+\beta(S=0)\right]\Big/dq^2dq_0$$
$$=\frac{G^2\cos^2\theta_c}{4\pi E^2}[q^2\alpha^{(\pm)}+(2E^2-2Eq_0-\tfrac{1}{2}q^2)\beta^{(\pm)}\mp(2E-q_0)q^2\gamma^{(\pm)}]. \quad (4)$$

The form factors $\alpha^{(\pm)}$, $\beta^{(\pm)}$, and $\gamma^{(\pm)}$ are functions of q^2 and q_0. The local commutation relations of Eqs. (1a) and (1c) imply that

$$2=\int_{0^-}^{\infty}dq_0[\beta^{(-)}-\beta^{(+)}]. \quad (5)$$

Combined with Eq. (4) and with the expression for

* Work supported in part by the U. S. Atomic Energy Commission. Prepared under Contract No. AT(11-1)-68 for the San Francisco Operations Office, U. S. Atomic Energy Commission.
† Junior Fellow, Society of Fellows, 1964–66.
‡ National Science Foundation Postdoctoral Fellow, 1965–1966.
[1] M. Gell-Mann, Physics 1, 63 (1964).
[2] S. L. Adler, Phys. Rev. 143, B1144 (1966).
[3] J. D. Bjorken, Phys. Rev. Letters 16, 408 (1966).

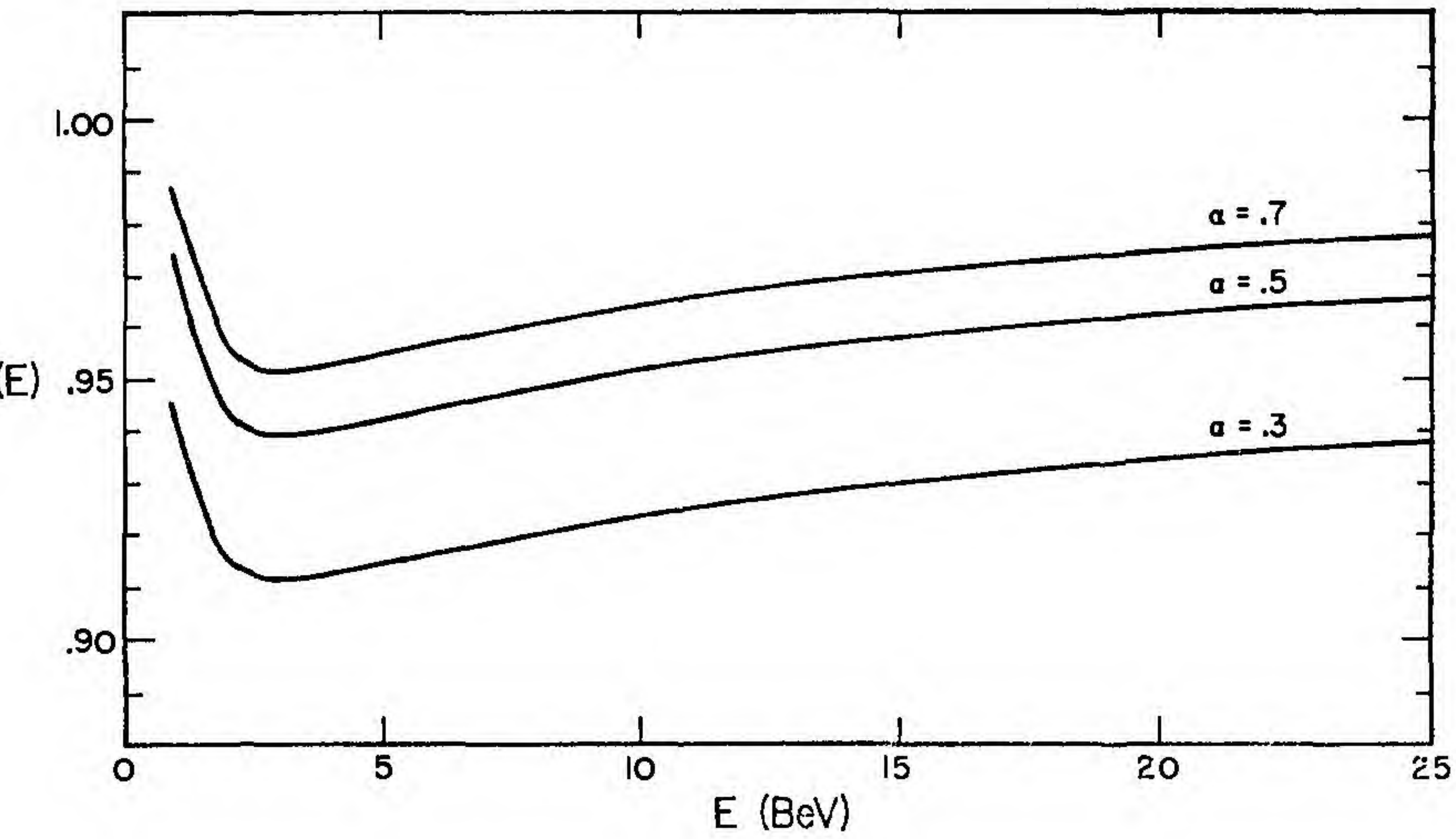


FIG. 1. The function $F_1(E)$ defined in Eq. (12). Above $\nu=5$ BeV, we have assumed $\sigma^{(-)}-\sigma^{(+)}\propto \nu^{-\alpha}$. See Ref. 7 for details of the numerical evaluation.

$d\sigma/dq^2$,

$$\frac{d\sigma}{dq^2}=\int_0^{E[1-q^2/(4E^2)]} dq_0\frac{d^2\sigma}{dq^2dq_0}, \tag{6}$$

Eq. (5) implies that

$$\lim_{E\to\infty}\left\{\frac{d\sigma_T[\bar{\nu}+p\to\beta(S=0)]}{dq^2}-\frac{d\sigma_T[\nu+p\to\beta(S=0)]}{dq^2}\right\}$$
$$=\frac{G^2\cos^2\theta_c}{\pi}. \tag{7}$$

In order to study the manner in which the limit in Eq. (7) is approached, let us separate $\beta^{(\pm)}$ in the sum rule of Eq. (5) into vector and axial-vector parts,

$$\beta^{(\pm)}=\beta_V^{(\pm)}+\beta_A^{(\pm)}, \tag{8}$$

which separately obey the sum rules[2]

$$1=\int_{0^-}^{\infty} dq_0[\beta_A^{(-)}-\beta_A^{(+)}]$$
$$=[g_A(q^2)]^2+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}$$
$$\times dW[\beta_A^{(-)}(q^2,W)-\beta_A^{(+)}(q^2,W)], \tag{9a}$$

$$1=\int_{0^-}^{\infty} dq_0[\beta_V^{(-)}-\beta_V^{(+)}]$$
$$=[F_1^V(q^2)]^2+q^2[F_2^V(q^2)]^2+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}$$
$$\times dW[\beta_V^{(-)}(q^2,W)-\beta_V^{(+)}(q^2,W)]. \tag{9b}$$

At $q^2=0$ and for $W\geq M_N+M_\pi$, $\beta_V^{(\pm)}(0,W)=0$, so that Eq. (9b) becomes the trivial equality $1=1$. Since

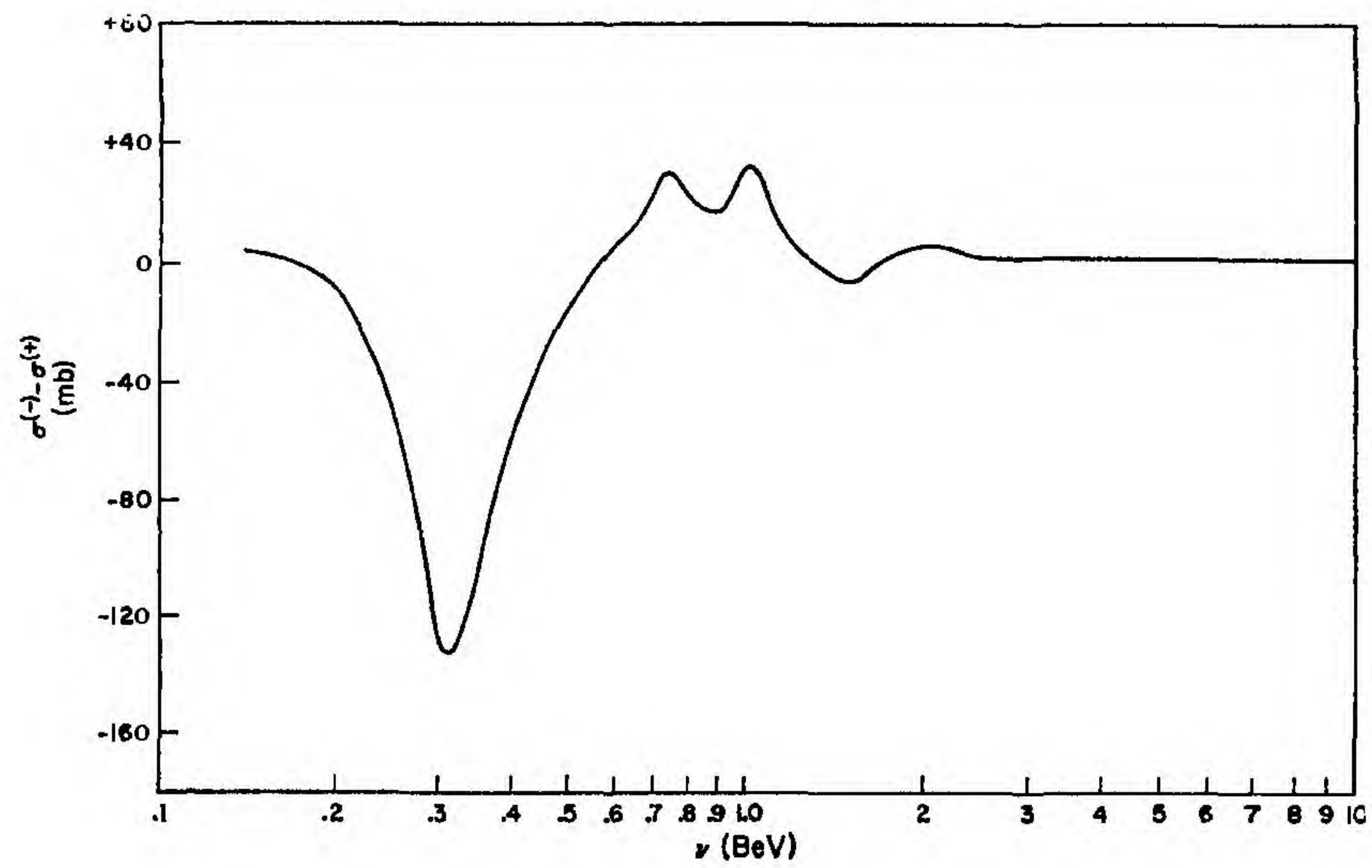


FIG. 2. Input values (Ref. 7) of $\sigma^{(-)}-\sigma^{(+)}$ for values of ν from threshold to 10 BeV, including the asymptotic tail above 5 BeV for the case $\alpha=0.5$.

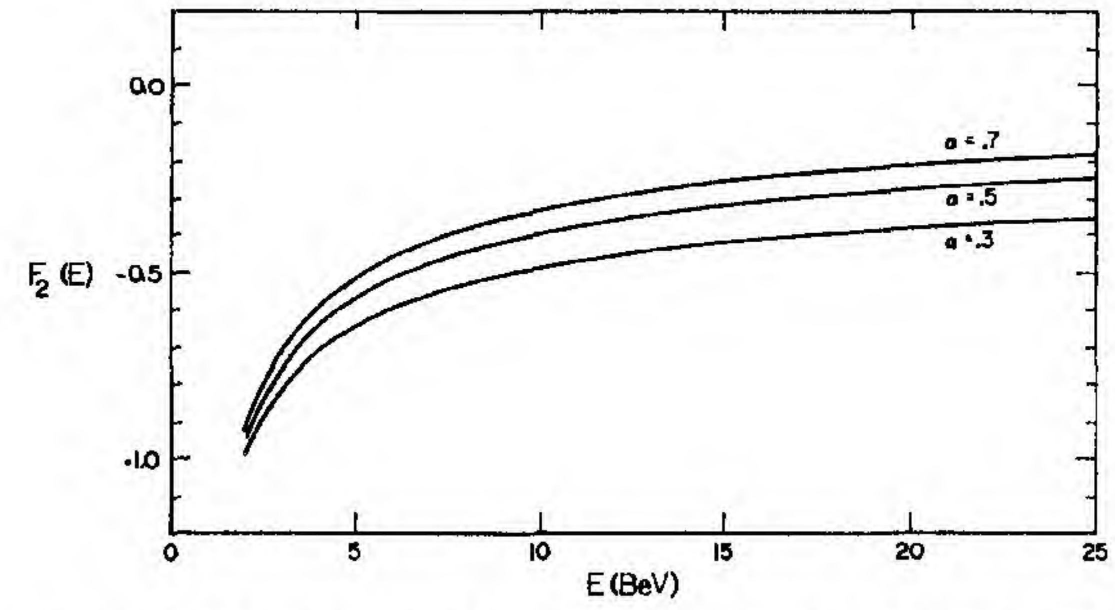


FIG. 3. The function $F_2(E)$ defined in Eq. (16). Above $q_0=1.1$ BeV, we have assumed

$2\sigma_T[\gamma(I=1)+p\rightarrow I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\rightarrow I=\frac{3}{2}]\propto q_0^{-\alpha}$.

See Ref. 10 for details of the numerical evaluation.

$\beta_A{}^{(\pm)}(0,W)$ is related to $\sigma^{(\pm)}(0,W)$, the total cross section for the scattering of a zero-mass $\pi^\pm$ on a proton, by[4]

$$\beta_A{}^{(\pm)}(0,W)=\frac{4M_N{}^3g_A{}^2}{\pi g_r(0)^2}\frac{\sigma^{(\pm)}(0,W)}{W^2-M_N{}^2},\qquad(10)$$

we see that Eq. (9a) becomes the usual sum rule for g_A.[5,6] Thus, Eqs. (4)–(7) imply that, at $q^2=0$,

$$\left\{\frac{d\sigma_T[\bar\nu+p\rightarrow\beta(S=0)]}{dq^2}-\frac{d\sigma_T[\nu+p\rightarrow\beta(S=0)]}{dq^2}\right\}\bigg|_{q^2=0}=\frac{G^2\cos^2\theta_c}{2\pi}[1+F_1(E)],\qquad(11)$$

with

$$F_1(E)=g_A{}^2\left\{1-\int_{M_\pi+M_\pi{}^2/(2M_N)}^{E}\frac{dq_0}{q_0}\left(1-\frac{q_0}{E}\right)\times\frac{2M_N{}^2}{\pi g_r(0)^2}[\sigma^{(+)}(0,W)-\sigma^{(-)}(0,W)]\right\}.\qquad(12)$$

In Eq. (12), $q_0=(W^2-M_N{}^2)/(2M_N)$, as obtained from Eq. (3) for q_0 with $q^2=0$. Rather than using Eq. (12) for our numerical analysis, we use the expression

$$F_1(E)\approx g_A{}^2\left\{1-\int_{M_\pi}^{E}\frac{d\nu}{\nu^2}(\nu^2-M_\pi{}^2)^{1/2}\left(1-\frac{\nu}{E}\right)\times\frac{2M_N{}^2}{\pi g_r{}^2}[\sigma^{(+)}(W)-\sigma^{(-)}(W)]\right\},\qquad(13)$$

which involves only the pion on-mass-shell cross sections $\sigma^{(\pm)}(W)\equiv\sigma^{(\pm)}(-M_\pi{}^2,W)$. The statement that the right-hand side of Eq. (13) approaches 1 as $E\rightarrow\infty$ is the polology form of the g_A sum rule.[5] The variable $\nu=(W^2-M_N{}^2-M_\pi{}^2)/(2M_N)$ is the pion laboratory energy. If PCAC is valid, the integrands of Eqs. (12) and (13) are expected to differ appreciably only for *small* center-of-mass energy W, where kinematical threshold effects may be important; this difference should not greatly affect the *large-E* behavior of $F_1(E)$.

The numerical evaluation[7] of $F_1(E)$ is shown in Fig. 1 [in Fig. 2 we plot the input data $\sigma^{(-)}-\sigma^{(+)}$]. It is seen that for energies of a few BeV, $F_1(E)$ becomes monotonic and greater than 90% of its asymptotic value of unity. Thus, neutrino energies of order 5 BeV are certainly adequate for testing the local current algebra at $q^2=0$.

The curve of Fig. 1 has no direct bearing on Bjorken's inequalities for electron scattering, since these inequalities come from the vector sum rule of Eq. (9b) and do not involve the axial-vector sum rule of Eq. (9a). As we remarked above, Eq. (9b) becomes a trivial identity at $q^2=0$. However, the first derivative of Eq. (9b) with respect to q^2 gives the interesting sum rule

$$0=2\frac{d}{dq^2}F_1{}^V(q^2)|_{q^2=0}+[F_2{}^V(0)]^2+\int_{M_N+M_\pi}^{\infty}\frac{W}{M_N}dW\frac{d}{dq^2}\times[\beta_V{}^{(-)}(q^2,W)-\beta_V{}^{(+)}(q^2,W)]|_{q^2=0},\qquad(14)$$

which has been derived by Cabibbo and Radicati and others.[2,8] To exploit this fact, let us keep only the vector part of Eqs. (4)–(7) and expand in a power series in q^2 {we use the fact that $\alpha_V{}^{(\pm)}|_{q^2=0}=[q_0{}^2d\beta_V{}^{(\pm)}/dq^2]|_{q^2=0}=V_1{}^{(\pm)}(0,W)$, in the notation of Ref. 2}:

$$\left\{\frac{d\sigma_{TV}[\bar\nu+p\rightarrow\beta(S=0)]}{dq^2}-\frac{d\sigma_{TV}[\nu+p\rightarrow\beta(S=0)]}{dq^2}\right\}=\frac{G^2\cos^2\theta_c}{2\pi}\left\{1+\frac{q^2}{M_N{}^2}F_2(E)+O[(q^2)^2]\right\},\qquad(15)$$

with

$$\frac{F_2(E)}{M_N{}^2}=2\frac{d}{dq^2}F_1{}^V(q^2)|_{q^2=0}+[F_2{}^V(0)]^2\frac{1}{2M_NE}\frac{1}{4E^2}+\int_{M_\pi+M_\pi{}^2/(2M_N)}^{E}\frac{dq_0}{q_0{}^2}\left[1-\frac{q_0}{E}+\frac{q_0{}^2}{2E^2}\right]\times[V_1{}^{(-)}(0,W)-V_1{}^{(+)}(0,W)],\qquad(16)$$

and $q_0=(W^2-M_N{}^2)/(2M_N)$ as in Eq. (12). The integrand in Eq. (16) may be related to the total cross

[4] In writing Eq. (9), we are *defining* the pion interpolating field to be the divergence of the axial-vector current, suitably normalized. The partially conserved axial-vector current (PCAC) hypothesis is used when we replace Eq. (12) by Eq. (13).

[5] W. I. Weisberger, Phys. Rev. Letters **14**, 1047 (1965).

[6] S. L. Adler, Phys. Rev. Letters **14**, 1051 (1965).

[7] We have used the cross-section tabulation of G. Höhler, C. Ebel, and J. Giesecke [Z. Physik **180**, 430 (1964)]. Above $\nu=5$ BeV, we have assumed $\sigma^{(+)}-\sigma^{(-)}=[\sigma^{(+)}-\sigma^{(-)}]|_{\nu=5\,\mathrm{BeV}}(\nu/5\text{ BeV})^{-\alpha}$. For each value of α, we have normalized F_1, so that $F_1(\infty)=1$.

[8] S. L. Adler (unpublished); J. D. Bjorken (unpublished); Phys. Rev. **148**, 1467 (1966); N. Cabibbo and L. Radicati, Phys. Letters **19**, 697 (1966); R. F. Dashen and M. Gell-Mann, in *Proceedings of the Third Coral Gables Conference on Symmetry Principles at High Energy* (W. H. Freeman and Company, San Francisco, 1966).

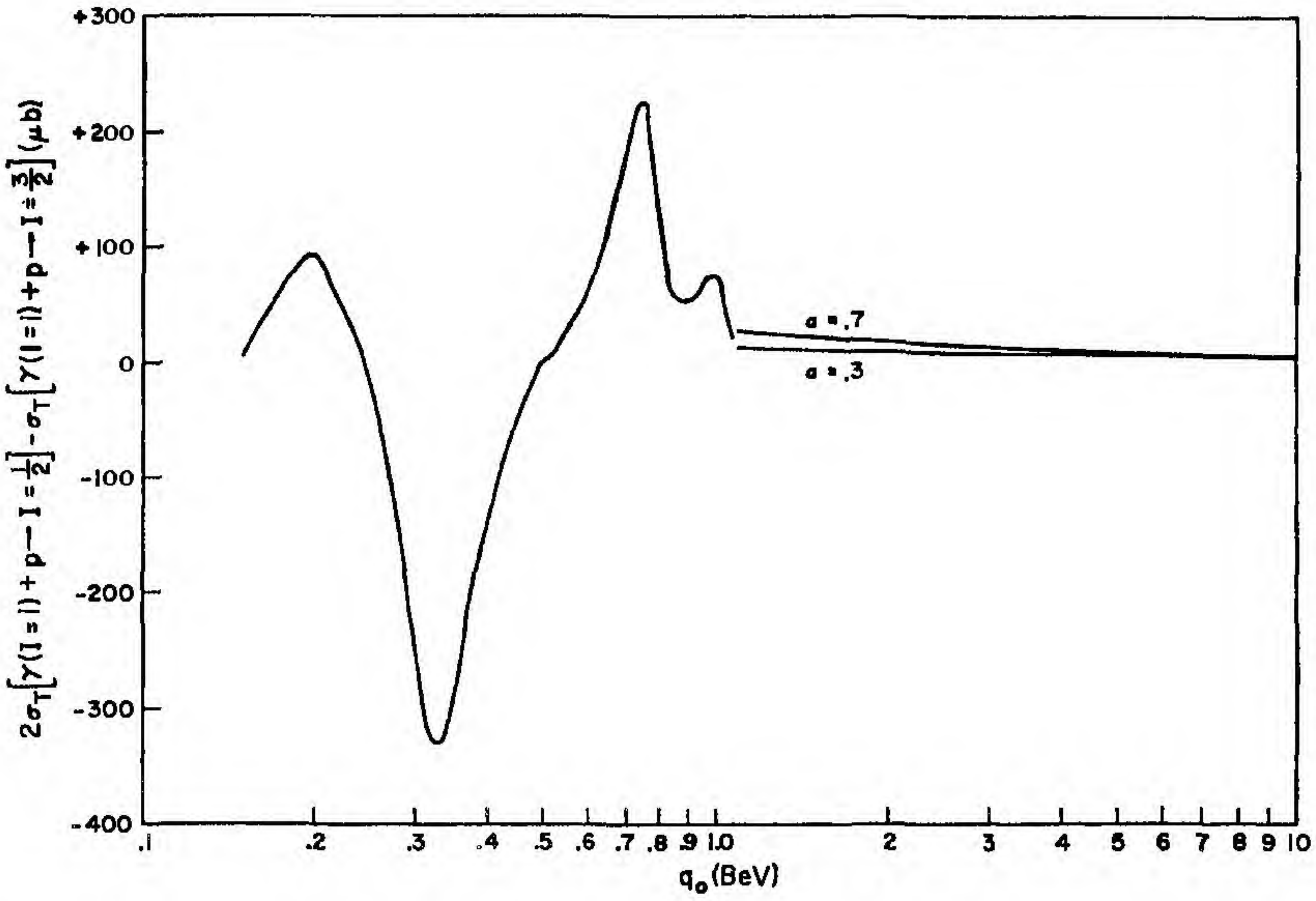


FIG. 4. Input values (Ref. 10) of $2\sigma_T[\gamma(I=1)+p\to I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\frac{3}{2}]$ for values of q_0 from threshold up to 10 BeV, including the asymptotic tail above $q_0=1.1$ BeV for the cases $\alpha=0.3$ and $\alpha=0.7$.

sections for *isovector* photons incident on nucleons to produce $I=\frac{1}{2}$ and $I=\frac{3}{2}$ final states:

$$V_1^{(-)}(0,W)-V_1^{(+)}(0,W)=\frac{q_0}{2\pi^2\alpha}\times\{2\sigma_T[\gamma(I=1)+p\to I=\tfrac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\tfrac{3}{2}]\}. \quad (17)$$

These isovector photon cross sections have been estimated by Gilman and Schnitzer,[9] who find that Eq. (14) appears to be satisfied. Using numerical estimates similar[10] to those of Gilman and Schnitzer, we have computed $F_2(E)$. The result, shown in Fig. 3, indicates that $|F_2(E)|$ is less than 0.5 for energies in the 5 to 10 BeV range. {In Fig. 4 and Table I we give the input data $2\sigma_T[\gamma(I=1)+p\to I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\frac{3}{2}]$.}

To conclude, for small q^2, energies of order 5 BeV suffice to test local commutation relations. We must caution that as q^2 increases, the needed energy $E(q^2,\delta)$ will be expected to increase rapidly. This is clear from the experimental fact that the single-nucleon and (3,3) resonance contributions, i.e., the small W contributions, to the sum rules of Eqs. (9a) and (9b) decrease rapidly with increasing q^2.[11] Thus, to maintain a constant sum

[9] F. J. Gilman and H. J. Schnitzer, Phys. Rev. **150**, 1362 (1966).

[10] Up to $q_0=1.1$ BeV, we have included the contributions of the s-wave, $N^*(1238)$, $N^*(1520)$, and $N^*(1688)$. We assumed the $N^*(1520)$ and $N^*(1688)$ peaks measured in photoproduction come only from isovector photon transitions, and that for these two resonances, $\Gamma_{\mathrm{elastic}}/\Gamma_{\mathrm{total}}=0.6$. Above 1.1 BeV, we have assumed $2\sigma_T[\gamma(I=1)+p\to I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\frac{3}{2}]=Nq_0^{-\alpha}$, and for each α we have chosen the normalization N of the tail to make $F_2(\infty)=0$. For none of the values of α considered did this require an unreasonably large tail.

[11] Nucleon form factors: E. B. Hughes *et al.*, Phys. Rev. **139**, B458 (1965); electroproduction: A. A. Cone *et al.*, Phys. Rev. Letters **14**, 326 (1965); weak production: CERN Report No. NPA/Int. 65-11, 1965 (unpublished).

TABLE I. Values of $2\sigma_T[\gamma(I=1)+p\to I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\frac{3}{2}]$ up to $q_0=1.1$ BeV.

q_0 (MeV)	$2\sigma_T[\gamma(I=1)+p\to I=\frac{1}{2}]-\sigma_T[\gamma(I=1)+p\to I=\frac{3}{2}]$ (μb)
150	+8
175	+60
200	+94
225	+50
250	−11
275	−108
300	−238
325	−328
350	−287
375	−200
400	−138
425	−86
450	−49
475	−24
500	+3
525	+7
550	+30
575	+46
600	+64
625	+100
650	+114
675	+148
700	+182
725	+217
750	+227
775	+176
800	+125
825	+80
850	+54
875	+53
900	+54
925	+54
950	+66
975	+76
1000	+75
1025	+72
1050	+46
1075	+22
1100	+10

at large q^2, the high W states, which require a large E to be excited, must make a much more important contribution to the sum rules than they do at $q^2=0$. The calculations of this paper shed no light on the important question of how rapidly $E(q^2,\delta)$ increases with q^2, but only serve to indicate at what energies E it may pay to begin the experimental study of the local current algebra.

PHYSICAL REVIEW VOLUME 151, NUMBER 4 25 NOVEMBER 1966

Low-Energy Theorem for the Weak Axial-Vector Vertex*

S. L. ADLER† AND Y. DOTHAN‡
California Institute of Technology, Pasadena, California
(Received 13 May 1966)

A low-energy theorem is derived for the weak axial-vector vertex. The theorem enables one to calculate from strong or electromagnetic processes the two leading terms in the expansion of the axial-vector vertex in powers of the leptonic four-momentum transfer. Applications to weak pion production, K_{e4} decay, and radiative μ capture are discussed. In particular, we express the radiative μ-capture matrix element, up to and including contributions linear in the leptonic four-momentum transfer and the photon four-momentum, in terms of the elastic weak form factors and pion photoproduction amplitudes.

INTRODUCTION

IT is well known[1] that the infrared divergent order k^{-1} term in the matrix element for the radiation of a photon of four-momentum k in any process (the matrix element of the electric current) can be expressed solely in terms of the matrix element for the same process with no current present. Low[2] has shown that current conservation enables one to calculate the electric-current matrix element not only to order k^{-1} but also to order k^0 in terms of the process without the current. In the present work, we derive analogous results for the matrix elements of the axial-vector current. We express each such matrix element in terms of the matrix element for the process with no axial-vector current and the matrix element of the *divergence* of the axial-vector current. The relation is exact to orders k^{-1} and k^0. Under the assumption of a partially conserved axial-vector current (PCAC),[3] we can relate the matrix element of the divergence to the corresponding matrix element of the pion source, which is physically measurable, apart from the usual small off-mass-shell extrapolation.[4] Thus we obtain an expression for the axial-vector matrix element solely in terms of physically measurable quantities. Clearly, this shows that the essential point in Low's derivation is not current conservation, but the fact that the divergence of the current is independently measurable. Results analogous to ours will hold for any current whose divergence is known.

In Sec. I we state two simple lemmas and rederive Low's results from them. In Sec. II we derive the analogous results for the strangeness-conserving weak axial-vector current. We also show how these results are modified when two currents are present, instead of only one. As an application, we treat in Sec. III the following processes: Weak pion production, K_{e4} decay, and radiative μ capture. In particular, we find in the case of radiative μ capture that when terms of order qk and higher are neglected (q=lepton four momentum transfer, k=photon four-momentum), the matrix element can be expressed solely in terms of the elastic weak form factors and pion photoproduction amplitudes. This means that structure effects linear in q or linear in k are determined, giving the leading corrections to the radiative μ capture matrix element previously calculated by Manacher and Wolfenstein[5] and by Opat.[6]

I. LOW'S RESULTS FOR THE ELECTROMAGNETIC CURRENT

We consider the process $a \to b+\gamma$, where a and b are arbitrary hadron states. The matrix element for the process is given by[7]

$$_{\rm out}\langle b\gamma | a\rangle_{\rm in} = ie(2\pi)^4\delta^{(4)}(p_a-p_b-k) \times\frac{1}{(2\pi)^{3/2}(2k_0)^{1/2}}N_aN_b\epsilon_\alpha^\star M_\alpha\,, \quad (1)$$

where p_a, p_b, N_a, and N_b are, respectively, the total four-momenta and the normalization factors of the particles in states a and b, ϵ_α is the polarization of the photon, and k is its four-momentum. The quantity M_α is related to the matrix element of the electromagnetic current $J_\alpha^{\rm EM}$ by

$$N_aN_bM_\alpha = {}_{\rm out}\langle b| J_\alpha^{\rm EM} | a\rangle_{\rm in}\,. \quad (2)$$

Conservation of the electromagnetic current implies that

$$k_\alpha M_\alpha = 0\,. \quad (3)$$

We state two simple mathematical lemmas from which Low's results are easily derived. [In the follow-

* Work supported in part by the U. S. Atomic Energy Commission. Prepared under Contract AT(11-1)-68 for the San Francisco Operations Office, U. S. Atomic Energy Commission.
† Junior Fellow, Society of Fellows. Present address: The Institute for Advanced Study, Princeton, New Jersey.
‡ On leave of absence from Israel Atomic Energy Commission, Soreq Research Establishment, Yavne, Israel.
[1] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Reading Massachusetts, 1955), p. 391.
[2] F. E. Low, Phys. Rev. **110**, 974 (1958).
[3] M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960); Y. Nambu, Phys. Rev. Letters **4**, 380 (1960); K. C. Chou, Zh. Eksperim. i Teor. Fiz. [English transl.: Soviet Phys.—JETP 12, 492 (1961)].
[4] S. L. Adler, Phys. Rev. **140**, B736 (1965), Sec. IIIC.
[5] G. K. Manacher and L. Wolfenstein, Phys. Rev. **116**, 782 (1959).
[6] G. I. Opat, Phys. Rev. **134**, B428 (1964).
[7] Four-vectors have an imaginary fourth component: $p=(\mathbf{p},p_4)=(\mathbf{p},ip_0)$ and $p\cdot q=\mathbf{p}\cdot\mathbf{q}+p_4q_4=\mathbf{p}\cdot\mathbf{q}-p_0q_0$. The quantity $p^\star$ is defined by $\mathbf{p}^\star=\mathbf{p}^*$, $p_4^\star=-p_4^*$, where * denotes the ordinary complex conjugate. The γ matrices ($\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5=\gamma_1\gamma_2\gamma_3\gamma_4$) are all Hermitian, and satisfy $\gamma_\alpha\gamma_\beta+\gamma_\beta\gamma_\alpha=2\delta_{\alpha\beta}$.

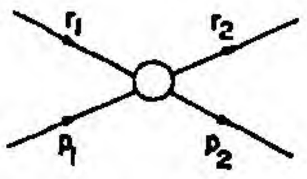


FIG. 1. The nonradiative process.

ing, $O(k^n)$ denotes terms of the nth or higher degree in k.]

Lemma 1: Let $M_\alpha^{\rm II}$ be an arbitrary four-vector function of arbitrary independent variables, which is independent of the four-vector k_α. Then $k_\alpha M_\alpha^{\rm II}=O(k^2)$ implies that $M_\alpha^{\rm II}=0$. Proof: Obvious.

Lemma 2: If $k_\alpha M_\alpha=0$ and $M_\alpha=M_\alpha^{\rm I}+M_\alpha^{\rm II}+O(k)$, where $M_\alpha^{\rm II}$ is independent of k and where $k_\alpha M_\alpha^{\rm I}=0$, then $M_\alpha=M_\alpha^{\rm I}+O(k)$. Proof: $k_\alpha M_\alpha=k_\alpha M_\alpha^{\rm I}=0$ implies $k_\alpha M_\alpha^{\rm II}=O(k^2)$, so by Lemma 1, $M_\alpha^{\rm II}=0$.

Note that "independent of k" is not the same as "zeroth order in k." For example, $k_\alpha/p\cdot k$ is zeroth order in k but is not independent of k.

We now apply the lemmas to the two cases considered by Low. First we discuss scattering of a charged scalar particle from a neutral scalar particle (Fig. 1). We denote the initial and final neutral-particle four-momenta by r_1 and r_2, and the corresponding charged-particle four-momenta by p_1 and p_2. Let $T(s=p_1\cdot r_1+p_2\cdot r_2,\ t=(r_1-r_2)^2,\ \Delta_1=p_1^2+M_1^2,\ \Delta_2=p_2^2+M_2^2)$ be the transition amplitude for the nonradiative process in Fig. 1. We have explicitly indicated the dependence of T on the amount by which the external charged particles are off the mass shell, since the amplitude for the process in which the photon is emitted from one of the external charged particle lines involves the off-mass-shell nonradiative amplitude. The physical nonradiative amplitude is $T(s,t,0,0)$.

The radiative amplitude gets contributions from two types of terms: terms in which the photon is radiated from an external charged particle line [Figs. 2(a) and 2(b); we call these terms $M_\alpha^{\rm ext}$] and terms in which the photon is radiated from an internal line [Fig. 2(c); we call these terms $M_\alpha^{\rm int}$]. The infrared divergent terms come only from $M_\alpha^{\rm ext}$, while $M_\alpha^{\rm int}$ is finite at $k=0$. We write

$$M_\alpha^{\rm int}(k)=M_\alpha^{\rm int}(0)+O(k). \tag{4}$$

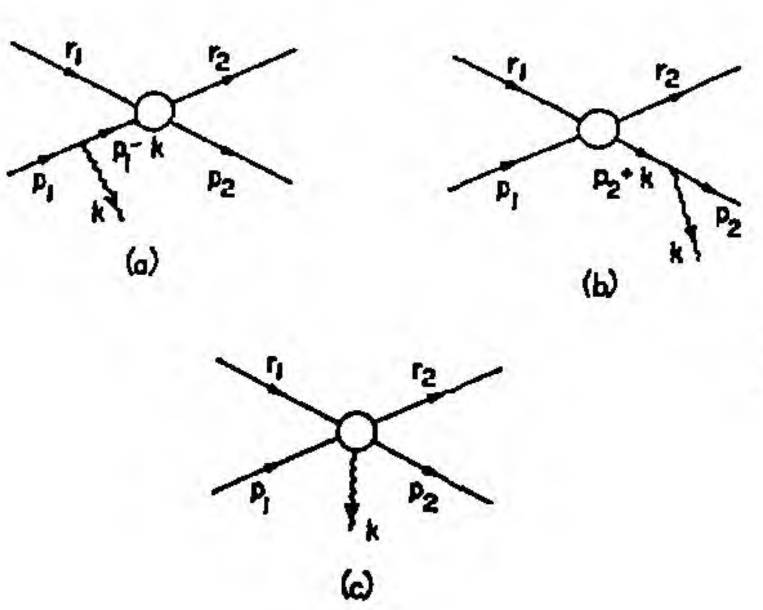


FIG. 2. Contributions to the radiative process.

We can express $M_\alpha^{\rm ext}$ in terms of T,

$$M_\alpha^{\rm ext}=\frac{(2p_2+k)_\alpha}{(p_2+k)^2+M_2^2}T[s+r_2\cdot k,t,0,(p_2+k)^2+M_2^2]+T[s-r_1\cdot k,t,(p_1-k)^2+M_1^2,0]\frac{(2p_1-k)_\alpha}{(p_1-k)^2+M_1^2}. \tag{5}$$

We expand T with respect to k, giving

$$\begin{aligned}M_\alpha^{\rm ext}=&\frac{(2p_2+k)_\alpha}{(2p_2+k)\cdot k}T[s,t,0,0]-T[s,t,0,0]\frac{(2p_1-k)_\alpha}{(2p_1-k)\cdot k}\\&+\left(\frac{p_{2\alpha}}{p_2\cdot k}r_2\cdot k+\frac{p_{1\alpha}}{p_1\cdot k}r_1\cdot k\right)\frac{\partial}{\partial s}T[s,t,0,0]\\&+2p_{2\alpha}\frac{\partial}{\partial\Delta_2}T[s,t,0,\Delta_2]\bigg|_{\Delta_2=0}\\&+2p_{1\alpha}\frac{\partial}{\partial\Delta_1}T[s,t,\Delta_1,0]\bigg|_{\Delta_1=0}+O(k).\end{aligned} \tag{6}$$

We are now able to rewrite M_α in the form required by Lemma 2,

$$M_\alpha=M_\alpha^{\rm ext}+M_\alpha^{\rm int}=M_\alpha^{\rm I}+M_\alpha^{\rm II}+O(k), \tag{7}$$

$$\begin{aligned}M_\alpha^{\rm I}=&\frac{(2p_2+k)_\alpha}{(2p_2+k)\cdot k}T[s,t,0,0]-T[s,t,0,0]\frac{(2p_1-k)_\alpha}{(2p_1-k)\cdot k}\\&+\left(\frac{p_{2\alpha}}{p_2\cdot k}r_2\cdot k+\frac{p_{1\alpha}}{p_1\cdot k}r_1\cdot k-r_{2\alpha}-r_{1\alpha}\right)\\&\times\frac{\partial}{\partial s}T[s,t,0,0],\end{aligned} \tag{7a}$$

$$\begin{aligned}M_\alpha^{\rm II}=&(r_{2\alpha}+r_{1\alpha})\frac{\partial}{\partial s}T[s,t,0,0]\\&+2p_{2\alpha}\frac{\partial}{\partial\Delta_2}T[s,t,0,\Delta_2]\bigg|_{\Delta_2=0}\\&+2p_{1\alpha}\frac{\partial}{\partial\Delta_1}T[s,t,\Delta_1,0]\bigg|_{\Delta_1=0}+M_\alpha^{\rm int}(0).\end{aligned} \tag{7b}$$

From this we conclude that $M_\alpha=M_\alpha^{\rm I}+O(k)$. In other words, the terms in the radiative amplitude of order k^0 as well as those of order k^{-1} have been determined.

The procedure required by the lemmas may be reduced to a simple recipe: (1) Write down $M_\alpha^{\rm ext}$, the sum of the terms in which the photon is radiated from an external charged particle line. (2) Drop all terms from $M_\alpha^{\rm ext}$ which are explicitly independent of k, giving a truncated amplitude $M_\alpha^{\rm ext\prime}$. (3) Add to $M_\alpha^{\rm ext\prime}$ a ΔM_α independent of k so as to make $k_\alpha(M_\alpha^{\rm ext\prime}+\Delta M_\alpha)=O(k^2)$. Then $M_\alpha^{\rm ext\prime}+\Delta M_\alpha$ is the $M_\alpha^{\rm I}$ required by the lemma.

Let us apply this recipe to the problem considered

above. We have computed M_α^{ext} in Eq. (5). In the first term let us expand T with respect to the off-mass-shell variable but not with respect to the energy variable:

$$T[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] =T[s+r_2\cdot k,\, t,\, 0,\, 0]+[(p_2+k)^2+M_2^2] \times\left\{\frac{\partial}{\partial\Delta_2}T[s+r_2\cdot k,\, t,\, 0,\, \Delta_2]\bigg|_{\Delta_2=0}+O(k)\right\}. \quad (8)$$

The off-mass-shell derivative term in this expansion, when substituted into Eq. (5), leads only to terms which are either explicitly independent of k or are of first order in k. These terms are dropped in forming the truncated matrix element. We repeat this procedure for the second term in Eq. (5). Thus the truncated matrix element $M_\alpha^{\text{ext}\prime}$ is

$$M_\alpha^{\text{ext}\prime}=\frac{(2p_2+k)_\alpha}{(2p_2+k)\cdot k}T[s+r_2\cdot k,\, t,\, 0,\, 0] -T[s-r_1\cdot k,\, t,\, 0,\, 0]\frac{(2p_1-k)_\alpha}{(2p_1-k)\cdot k}+O(k). \quad (9)$$

The divergence of $M_\alpha^{\text{ext}\prime}$ is

$$k_\alpha M_\alpha^{\text{ext}\prime}=T[s+r_2\cdot k,\, t,\, 0,\, 0] -T[s-r_1\cdot k,\, t,\, 0,\, 0]+O(k^2) =(r_2\cdot k+r_1\cdot k)\frac{\partial}{\partial s}T[s,t,0,0]+O(k^2). \quad (10)$$

Hence, ΔM_α is determined to be

$$\Delta M_\alpha=-(r_2+r_1)_\alpha\frac{\partial}{\partial s}T[s,t,0,0]. \quad (11)$$

Clearly, $M_\alpha^{\text{ext}\prime}+\Delta M_\alpha$ is identical with the M_α^{I} of Eq. (7a) to order k.

As a second illustration of the procedure, we consider the case when the charged particles have spin $\frac{1}{2}$. This is the simplest photon analog of the axial-vector case, since the axial-vector vertex cannot couple to a spin-zero particle line. As we shall see, the only difference from the preceding case is due to slight complications caused by spin.

We start by writing down M_α^{ext},

$$M_\alpha^{\text{ext}}=\bar{u}(p_2)\left\{(i\gamma_\alpha+i\frac{\mu}{2M_2}\sigma_{\alpha\beta}k_\beta)\frac{1}{i\gamma\cdot(p_2+k)+M_2} \times T[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] +T[s-r_1\cdot k,\, t,\, (p_1-k)^2+M_1^2,\, 0] \times\frac{1}{i\gamma\cdot(p_1-k)+M_1}\left(i\gamma_\alpha+i\frac{\mu}{2M_1}\sigma_{\alpha\beta}k_\beta\right)\right\}u(p_1). \quad (12)$$

Let us discuss the first term of Eq. (12). Because the final fermion is off its mass shell, $T[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2]$ contains terms which give a vanishing contribution as $k\to 0$ when multiplied on the left by a spinor $\bar{u}(p_2)$. These terms are not physically measurable in the nonradiative process. It is therefore convenient to write T in the form

$$T[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] =\frac{i\gamma\cdot(p_2+k)+W}{2W}T^N[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] +\frac{-i\gamma\cdot(p_2+k)+W}{2W}T^P[s+r_2\cdot k,\, t,\, 0, \times(p_2+k)^2+M_2^2], \quad (13)$$

where W denotes $[-(p_2+k)^2]^{1/2}$. The term $T^P[s,t,0,0]$ is the amplitude measured in the nonradiative process. We rearrange Eq. (13) in the form

$$T[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] =T^P[s+r_2\cdot k,\, t,\, 0,\, 0]+[i\gamma\cdot(p_2+k)+M_2] \times\Big\{[-i\gamma\cdot(p_2+k)+M_2]\frac{\partial}{\partial\Delta_2} \times T^P[s+r_2\cdot k,\, t,\, 0,\, \Delta_2]\bigg|_{\Delta_2=0}+O(k) +\frac{1}{2W}\left[1+\frac{i\gamma\cdot(p_2+k)-M_2}{W+M_2}\right] \times\{T^N[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2] -T^P[s+r_2\cdot k,\, t,\, 0,\, (p_2+k)^2+M_2^2]\}\Big\}. \quad (14)$$

When substituted back into Eq. (12), the term in boldface brackets in Eq. (14) leads to terms either independent of k or of first order in k. Strictly speaking, we should have included in Eq. (12) the negative-frequency terms in the photon–spin-$\frac{1}{2}$–off-mass-shell spin-$\frac{1}{2}$ vertex. By the same argument, these terms do not contribute to the truncated matrix element. Hence, the truncated matrix element is

$$M_\alpha^{\text{ext}\prime}=\bar{u}(p_2)\left\{\left(i\gamma_\alpha+i\frac{\mu}{2M_2}\sigma_{\alpha\beta}k_\beta\right) \times\frac{1}{i\gamma\cdot(p_2+k)+M_2}T^P[s+r_2\cdot k,\, t,\, 0,\, 0] +T^P[s-r_1\cdot k,\, t,\, 0,\, 0]\frac{1}{i\gamma\cdot(p_1-k)+M_1} \times\left(i\gamma_\alpha+i\frac{\mu}{2M_1}\sigma_{\alpha\beta}k_\beta\right)\right\}u(p_1)+O(k), \quad (15)$$

which involves only the physically measurable matrix element. Using the identities

$$\bar{u}(p_2)i\gamma\cdot k\frac{1}{i\gamma\cdot(p_2+k)+M_2}=\bar{u}(p_2)\,,$$
$$\frac{1}{i\gamma\cdot(p_1-k)+M_1}i\gamma\cdot ku(p_1)=-u(p_1)\,, \tag{16}$$

we can calculate $k_\alpha M_\alpha{}^{\text{ext}\prime}$,

$$k_\alpha M_\alpha{}^{\text{ext}\prime}=\bar{u}(p_2)\{T^P[s+r_2\cdot k,t,0,0]-T^P[s-r_1\cdot k,t,0,0]\}u(p_1)+O(k^2)\,. \tag{17}$$

The expression between the spinors is identical to Eq. (10) in the spin-zero case. Therefore, ΔM_α is

$$\Delta M_\alpha=-(r_2+r_1)_\alpha\bar{u}(p_2)\frac{\partial}{\partial s}T^P[s,t,0,0]u(p_1)\,, \tag{18}$$

and $M_\alpha{}^{\mathrm{I}}$ is $M_\alpha{}^{\text{ext}\prime}+\Delta M_\alpha$. This is Low's result.

II. AXIAL-VECTOR CURRENT

We now consider the matrix element of the strangeness-conserving weak axial-vector current $J_\alpha{}^{Aj}$ between hadron states a and b,

$$N_aN_bM_\alpha{}^j={}_{\text{out}}\langle b|J_\alpha{}^{Aj}|a\rangle_{\text{in}}\,. \tag{19}$$

The superscript j is an isotopic spin index ($j=1,2,3$). We no longer have the equation $k_\alpha M_\alpha{}^j=0$, since the axial-vector current is not conserved. Let D^j be the matrix element of the divergence of the axial-vector current,

$$N_aN_bD^j=N_aN_bk_\alpha M_\alpha{}^j={}_{\text{out}}\langle b|-i\partial_\alpha J_\alpha{}^{Aj}|a\rangle_{\text{in}}\,. \tag{20}$$

Here, as in Section I, $k=p_a-p_b$. The PCAC hypothesis relates matrix elements of the divergence of the axial-vector current to matrix elements of the pion source,

$${}_{\text{out}}\langle b|\partial_\alpha J_\alpha{}^{Aj}|a\rangle_{\text{in}}=\frac{M_Ng_A}{g_r(0)}\frac{m_\pi{}^2}{k^2+m_\pi{}^2}{}_{\text{out}}\langle b|J_\pi{}^j|a\rangle_{\text{in}}\,, \tag{21}$$

where M_N and m_π are the nucleon and pion masses, $J_\pi{}^j$ is the pion source, $g_A\equiv g_A(0)\approx1.18$ is the weak axial-vector coupling constant, and $g_r(0)$ is the off-mass-shell pion-nucleon coupling constant. The [physical coupling constant is $g_r\equiv g_r(-m_\pi{}^2)$; $g_r{}^2/4\pi\approx14.6$.] We wish to emphasize that the PCAC hypothesis allows one to measure D^j in purely strong interaction experiments.

Since the axial-vector current is not conserved, we will need a slightly modified version of Lemma 2:

Lemma 2′: If $k_\alpha M_\alpha{}^j=D^j$ and $M_\alpha{}^j=M_\alpha{}^{j\mathrm{I}}+M_\alpha{}^{j\mathrm{II}}+O(k)$, where $M_\alpha{}^{j\mathrm{II}}$ is independent of k and where $k_\alpha M_\alpha{}^{j\mathrm{I}}=D^j+O(k^2)$, then $M_\alpha{}^j=M_\alpha{}^{j\mathrm{I}}+O(k)$. This lemma leads to a modification of the recipe stated in Sec. I: (1) Write down $M_\alpha{}^{j\,\text{ext}}$, the sum of terms in which the axial-vector current is coupled to external particle lines. (2) Drop all terms from $M_\alpha{}^{j\,\text{ext}}$ which are explicitly independent of k, giving a truncated amplitude $M_\alpha{}^{j\,\text{ext}\prime}$. (3) Add to $M_\alpha{}^{j\,\text{ext}\prime}$ a $\Delta M_\alpha{}^j$ independent of k so as to make $k_\alpha(M_\alpha{}^{j\,\text{ext}\prime}+\Delta M_\alpha{}^j)=D^j+O(k^2)$. Then $M_\alpha{}^{j\,\text{ext}\prime}+\Delta M_\alpha{}^j$ is the $M_\alpha{}^{\mathrm{I}}$ required by the lemma. We actually will not omit *all* terms of order k, but will consistently retain terms of order k which explicitly contain a pion propagator.

As an illustration of the recipe, we will consider the problem analogous to the second example in Sec. I, scattering of a spin-zero particle from a spin-$\frac{1}{2}$ particle (which we will take to be a nucleon) with an additional coupling of the spin-$\frac{1}{2}$ particle to the axial-vector current. The answer will involve the corresponding matrix element, in which the axial-vector current is replaced by the pion source. We write the pion-emission matrix element in the form

$$\begin{aligned}M_\pi{}^j&={}_{\text{out}}\langle b|J_\pi{}^j|a\rangle_{\text{in}}(N_aN_b)^{-1}\\&=\bar{u}(p_2)\Big\{ig_r(k^2)\tau^j\gamma_5\frac{1}{i\gamma\cdot(p_2+k)+M_N}\\&\quad\times T^P[s+r_2\cdot k,t,0,0]+T^P[s-r_1\cdot k,t,0,0]\\&\quad\times\frac{1}{i\gamma\cdot(p_1-k)+M_N}ig_r(k^2)\tau^j\gamma_5+i\bar{T}_\pi{}^j(0)\\&\quad+ik_\lambda\frac{\partial}{\partial k_\lambda}\bar{T}_\pi{}^j(k)\Big|_{k=0}+O(k^2)\Big\}u(p_1)\,.\end{aligned} \tag{22}$$

We have explicitly exhibited the Born terms in the form given by dispersion theory, where residues are evaluated at the Born pole and so no nucleon-off-mass-shell terms are present. The way we write the Born terms serves as the definition of the non-Born part $\bar{T}_\pi{}^j(k)$.

We are now ready to write down $M_\alpha{}^{j\,\text{ext}}$,

$$\begin{aligned}M_\alpha{}^{j\,\text{ext}}&=\bar{u}(p_2)\Big\{ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\frac{1}{i\gamma\cdot(p_2+k)+M_N}\\&\quad\times T[s+r_2\cdot k,t,0,(p_2+k)^2+M_N{}^2]\\&\quad+T[s-r_1\cdot k,t,(p_1-k)^2+M_N{}^2,0]\\&\quad\times\frac{1}{i\gamma\cdot(p_1-k)+M_N}ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\Big\}u(p_1)\\&\quad+\frac{M_Ng_A}{g_r(0)}\frac{ik_\alpha}{k^2+m_\pi{}^2}M_\pi{}^j.\end{aligned} \tag{23}$$

The term in brackets in Eq. (23) is the direct coupling of the axial-vector current to the external nucleon lines. The term proportional to $M_\pi{}^j$ comes from the diagrams shown in Fig. 3; although this term is formally of first order in k, it can be important because of the small mass of the pion.

As we have seen in Sec. I, the truncated matrix element is obtained by dropping the negative frequency part of T and by neglecting off-mass-shell terms. This gives

$$
\begin{aligned}
M_\alpha{}^{j\,\mathrm{ext}'}=\bar{u}(p_2)\Big\{ & ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\frac{1}{i\gamma\cdot(p_2+k)+M_N} \\
& \times T^P[s+r_2\cdot k,t,0,0]+T^P[s-r_1\cdot k,t,0,0] \\
& \times\frac{1}{i\gamma\cdot(p_1-k)+M_N}ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\Big\}u(p_1) \\
& +\frac{M_Ng_A}{g_r(0)}\frac{ik_\alpha}{k^2+m_\pi{}^2}M_\pi{}^j+O(k). \qquad (24)
\end{aligned}
$$

Using the identities

$$
\begin{aligned}
& \bar{u}(p_2)i\gamma\cdot k\gamma_5\frac{1}{i\gamma\cdot(p_2+k)+M_N} \\
& \quad =\bar{u}(p_2)\Big[-\gamma_5+2M_N\gamma_5\frac{1}{i\gamma\cdot(p_2+k)+M_N}\Big], \\
& \frac{1}{i\gamma\cdot(p_1-k)+M_N}i\gamma\cdot k\gamma_5u(p_1) \\
& \quad =\Big[-\gamma_5+\frac{1}{i\gamma\cdot(p_1-k)+M_N}2M_N\gamma_5\Big]u(p_1),
\end{aligned} \qquad (25)
$$

we can calculate $k_\alpha M_\alpha{}^{j\,\mathrm{ext}'}$,

$$
\begin{aligned}
k_\alpha M_\alpha{}^{j\,\mathrm{ext}'}=\bar{u}(p_2)\Big\{ & -\tfrac{1}{2}g_A\tau^j\gamma_5T^P[s+r_2\cdot k,t,0,0] \\
& -T^P[s-r_1\cdot k,t,0,0]\tfrac{1}{2}g_A\tau^j\gamma_5 \\
& +\frac{M_Ng_Am_\pi{}^2}{k^2+m_\pi{}^2}\Big\{\tau^j\gamma_5\frac{1}{i\gamma\cdot(p_2+k)+M_N} \\
& \times T^P[s+r_2\cdot k,t,0,0]+T^P[s-r_1\cdot k,t,0,0] \\
& \times\frac{1}{i\gamma\cdot(p_1-k)+M_N}\tau^j\gamma_5\Big\}+O(k^2)\Big\}u(p_1) \\
& -\frac{M_Ng_A}{g_r(0)}\frac{k^2}{k^2+m_\pi{}^2}\bar{u}(p_2)\Big\{\bar{T}_\pi{}^j(0)+k_\lambda\frac{\partial}{\partial k_\lambda} \\
& \times\bar{T}_\pi{}^j(k)\Big|_{k=0}+O(k^2)\Big\}u(p_1). \qquad (26)
\end{aligned}
$$

In deriving Eq. (26), we have combined the Born terms in $M_\pi{}^j$ with the divergence of the first term in Eq. (24), and have expanded the form factors $g_A(k^2)$ and $g_r(k^2)$ in powers of k^2.

We determine $\Delta M_\alpha{}^j$ by the requirement that

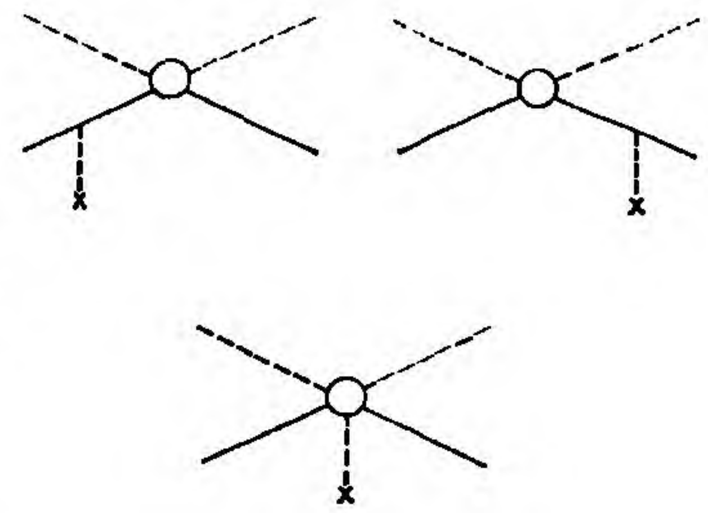


FIG. 3. Pion pole contributions to the axial-vector current matrix element. The axial-vector coupling is denoted by $\times$.

$k_\alpha(M_\alpha{}^{j\,\mathrm{ext}'}+\Delta M_\alpha{}^j)=D^j$, with

$$
\begin{aligned}
D^j=-i\frac{M_Ng_A}{g_r(0)}\frac{m_\pi{}^2}{k^2+m_\pi{}^2}\bar{u}(p_2) & \\
\times\Big\{ & ig_r(0)\tau^j\gamma_5\frac{1}{i\gamma\cdot(p_2+k)+M_N} \\
& \times T^P[s+r_2\cdot k,t,0,0]+T^P[s-r_1\cdot k,t,0,0] \\
& \times\frac{1}{i\gamma\cdot(p_1-k)+M_N}ig_r(0)\tau^j\gamma_5+i\bar{T}_\pi{}^j(0) \\
& +ik_\lambda\frac{\partial}{\partial k_\lambda}\bar{T}_\pi{}^j(k)\Big|_{k=0}+O(k^2)\Big\}u(p_1). \qquad (27)
\end{aligned}
$$

Comparing Eqs. (26) and (27), we see that $k_\alpha\Delta M_\alpha{}^j$ must satisfy

$$
\begin{aligned}
k_\alpha\Delta M_\alpha{}^j=\bar{u}(p_2)\Big\{ & \tfrac{1}{2}g_A\tau^j\gamma_5T^P[s+r_2\cdot k,t,0,0] \\
& +T^P[s-r_1\cdot k,t,0,0]\tfrac{1}{2}g_A\tau^j\gamma_5+\frac{M_Ng_A}{g_r(0)} \\
& \times\Big[\bar{T}_\pi{}^j(0)+k_\lambda\frac{\partial}{\partial k_\lambda}\bar{T}_\pi{}^j(k)\Big|_{k=0}\Big]+O(k^2)\Big\}u(p_1) \qquad (28a) \\
=\bar{u}(p_2)\Big\{ & \tfrac{1}{2}g_A\tau^j\gamma_5T^P[s,t,0,0] \\
& +T^P[s,t,0,0]\tfrac{1}{2}g_A\tau^j\gamma_5+\frac{M_Ng_A}{g_r(0)}\bar{T}_\pi{}^j(0)\Big\}u(p_1) \\
+k_\alpha\bar{u}(p_2)\Big\{ & r_{2\alpha}\tfrac{1}{2}g_A\tau^j\gamma_5\frac{\partial}{\partial s}T^P[s,t,0,0] \\
& -\frac{\partial}{\partial s}T^P[s,t,0,0]\tfrac{1}{2}g_A\tau^j\gamma_5r_{1\alpha} \\
& +\frac{M_Ng_A}{g_r(0)}\frac{\partial}{\partial k_\alpha}\bar{T}_\pi{}^j(k)\Big|_{k=0}\Big\}u(p_1)+O(k^2). \qquad (28b)
\end{aligned}
$$

As the reader has undoubtedly noted, the nucleon propagator terms have exactly cancelled between Eq. (26) and Eq. (27), and so do not appear in Eq. (28a). In the term involving $\bar{T}_\pi{}^j$, the pion propagator has dropped

out altogether, since

$$\frac{k^2}{k^2+m_\pi^2}+\frac{m_\pi^2}{k^2+m_\pi^2}=1. \tag{29}$$

In going from Eq. (28a) to Eq. (28b), we have simply expanded in powers of k and collected together the terms of zeroth, first, and second order in k.

Since $k_\alpha \Delta M_\alpha{}^j$ is of first order in k, the zeroth-order terms on the right-hand side of Eq. (28b) must vanish identically. This gives

$$\bar{u}(p_2)\bar{T}_\pi{}^j(0)u(p_1)=-\bar{u}(p_2)\left\{\frac{g_r(0)}{2M_N}\tau^j\gamma_5 T^P[s,t,0,0] + T^P[s,t,0,0]\frac{g_r(0)}{2M_N}\tau^j\gamma_5\right\}u(p_1). \tag{30}$$

This formula, which has been obtained previously,[8] expresses the matrix element for the emission of a zero four-momentum pion in terms of the matrix element of the process without the pion. Equation (30) can be used to eliminate $\bar{T}_\pi{}^j(0)$ from the term proportional to $M_\pi{}^j$ in Eq. (24). Comparing the terms of first order in k, we find

$$\Delta M_\alpha{}^j=\bar{u}(p_2)\left\{r_{2\alpha}\tfrac{1}{2}g_A\tau^j\gamma_5\frac{\partial}{\partial s}T^P[s,t,0,0] - \frac{\partial}{\partial s}T^P[s,t,0,0]\tfrac{1}{2}g_A\tau^j\gamma_5 r_{1\alpha} + \frac{M_N g_A}{g_r(0)}\frac{\partial}{\partial k_\alpha}\bar{T}_\pi{}^j(k)\bigg|_{k=0}\right\}u(p_1). \tag{31}$$

Adding this expression to the $M_\alpha{}^{j\,\mathrm{ext}'}$ of Eq. (24) gives the analog of Low's result for the axial-vector case.

A similar method can be applied to the case in which more than one current is acting. As an example, we consider the matrix element[9]

$$M_{\alpha\sigma}{}^j=\int d^4y\, e^{iq\cdot y}\,{}_{\mathrm{out}}\langle b|T[J_\alpha{}^{Aj}(x)J_\sigma(y)]|a\rangle_{\mathrm{in}}. \tag{32a}$$

Calculating $k_\alpha M_{\alpha\sigma}{}^j$, we get

$$\begin{aligned} k_\alpha M_{\alpha\sigma}{}^j &= \int d^4y e^{iq\cdot y}\left\langle b\left|-i\frac{\partial}{\partial x_\alpha}T[J_\alpha{}^{Aj}(x)J_\sigma(y)]\right|a\right\rangle_{\mathrm{in}}^{\mathrm{out}} \\ &= \int d^4y e^{iq\cdot y}\,{}_{\mathrm{out}}\langle b|-\delta(x_0-y_0)[J_4{}^{Aj}(x)J_\sigma(y)]|a\rangle_{\mathrm{in}} \\ &\quad+\int d^4y e^{iq\cdot y}\,{}_{\mathrm{out}}\langle b|-iT[\partial_\alpha J_\alpha{}^{Aj}(x)J_\sigma(y)]|a\rangle_{\mathrm{in}}. \end{aligned} \tag{32b}$$

The only difference from the case treated above is that the divergence, in addition to having the term with a pion vertex substituted for the axial-vector vertex, also contains an equal-time commutator term. Following the procedure of this section, we can determine $M_{\alpha\sigma}{}^j$, apart from terms of order k and higher. If the divergence of J_σ is also known, we can apply the technique a second time, determining terms of order k which are independent of q. This leaves an error which only involves terms of order qk and higher.[10] We will consider such a case in the next section, when we discuss radiative μ capture.

III. APPLICATIONS

In this section we apply the results of the previous section to several concrete examples. We consider first single-pion production from a nucleon by the axial-vector current. As an illustration of the use of our method in the strangeness-changing case, we discuss K_{e4} decay. We finally discuss the process of radiative μ capture on a proton, an example in which two currents are present.

1. Weak Pion Production

We consider the process

$$\nu(k_\nu)+N(p_1)\rightarrow l(k_l)+N(p_2)+\pi^n(q), \tag{33}$$

where the four-momentum of each particle is indicated in parentheses. Let $M_\alpha{}^{jn}$ be the axial-vector matrix element for this process, as defined in Eq. (19), with

$$\begin{aligned} |a\rangle_{\mathrm{in}} &= |N(p_1)\rangle, \\ {}_{\mathrm{out}}\langle b| &= {}_{\mathrm{out}}\langle N(p_2)\pi^n(q)|, \\ k &= k_l-k_\nu=p_1-(p_2+q). \end{aligned} \tag{34}$$

In this case, $T^P[s,t,0,0]$ is the pion-nucleon vertex $ig_r\gamma_5\tau^n$, which has no s dependence. Hence the $\partial/\partial s$ terms in Eq. (31) vanish. Clearly $M_\pi{}^{jn}$, the matrix element with the pion source substituted for the axial-vector current, is just the amplitude for pion-nucleon scattering. We find

$$\begin{aligned} M_\alpha{}^{jn\,\mathrm{ext}'} = \bar{u}(p_2)\bigg\{ & ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\frac{1}{i\gamma\cdot(p_2+k)+M_N} \\ & \times ig_r\gamma_5\tau^n + ig_r\gamma_5\tau^n\frac{1}{i\gamma\cdot(p_1-k)+M_N} \\ & \times ig_A(k^2)\gamma_\alpha\gamma_5\frac{\tau^j}{2}\bigg\}u(p_1) \\ & +\frac{M_N g_A}{g_r(0)}\frac{ik_\alpha}{k^2+m_\pi^2}M_\pi{}^{jn}, \end{aligned} \tag{35a}$$

[8] Y. Nambu and D. Lurié, Phys. Rev. 125, 1429 (1962); S. L. Adler, *ibid.* 139, B1638 (1965).

[9] In Eq. (32) we have neglected "seagull" terms, which will be included in the calculations of Sec. III.

[10] This method has been applied to the case when only vector currents are present by G. K. Manacher, thesis, Carnegie Institute of Technology Report NYO 9284, 1961 (unpublished).

$$
\begin{aligned}
M_\pi{}^{jn} = \bar{u}(p_2)\bigg\{ & ig_r(k^2)\tau^j\gamma_5 \frac{1}{i\gamma\cdot(p_2+k)+M_N} ig_r\gamma_5\tau^n \\
& + ig_r\gamma_5\tau^n \frac{1}{i\gamma\cdot(p_1-k)+M_N} ig_r(k^2)\tau^j\gamma_5 \\
& + i\bar{T}_\pi{}^{jn}(0) + ik_\lambda \frac{\partial}{\partial k_\lambda}\bar{T}_\pi{}^{jn}(k)\bigg|_{k=0} + O(k^2)\bigg\} u(p_1).
\end{aligned}
\tag{35b}
$$

From Eq. (30), we find that

$$
\begin{aligned}
& \bar{u}(p_2) i\bar{T}_\pi{}^{jn}(0) u(p_1) \\
& = -i\bar{u}(p_2)\bigg\{ \frac{g_r(0)}{2M_N}\tau^j\gamma_5 ig_r\gamma_5\tau^n + ig_r\gamma_5\tau^n \frac{g_r(0)}{2M_N}\tau^j\gamma_5 \bigg\} u(p_1) \\
& = \bar{u}(p_2)\bigg\{ \frac{g_r g_r(0)}{M_N}\delta^{nj} \bigg\} u(p_1).
\end{aligned}
\tag{36}
$$

From Eq. (31), we have

$$
\Delta M_\alpha{}^{jn} = \frac{M_N g_A}{g_r(0)} \bar{u}(p_2)\bigg\{ \frac{\partial}{\partial k_\alpha}\bar{T}_\pi{}^{jn}(k)\bigg|_{k=0} \bigg\} u(p_1). \tag{37}
$$

From the usual expression for the pion-nucleon scattering amplitude,[11] we find (remembering that $-k$ is the incoming pion four-momentum),

$$
\begin{aligned}
& \bar{u}(p_2)\bigg\{ \frac{\partial}{\partial k_\alpha}\bar{T}_\pi{}^{jn}(k)\bigg|_{k=0} \bigg\} u(p_1) \\
& = i\frac{\partial}{\partial k_\alpha}\bar{u}(p_2)\bigg\{ \bigg[-A^{\pi N(+)}\bigg(\nu = \frac{k\cdot(p_1+p_2)}{2M_N}, \\
& \qquad \nu_B = -\frac{q\cdot k}{2M_N}, k^2 \bigg) - i\gamma\cdot k\bar{B}^{\pi N(+)}(\nu,\nu_B,k^2) \bigg]\delta^{nj} \\
& \qquad + [-\bar{A}^{\pi N(-)}(\nu,\nu_B,k^2) - i\gamma\cdot k\bar{B}^{\pi N(-)}(\nu,\nu_B,k^2)] \\
& \qquad \times \tfrac{1}{2}[\tau^n,\tau^j] \bigg\} u(p_1)\,|_{k=0} \\
& = i\bar{u}(p_2)\bigg\{ \bigg[\frac{\partial\bar{A}^{\pi N(+)}}{\partial\nu_B}\bigg|_{\nu=\nu_B=k^2=0} \frac{q_\alpha}{2M_N} \bigg]\delta^{nj} \\
& \qquad - \bigg[\frac{\partial\bar{A}^{\pi N(-)}}{\partial\nu}\bigg|_{\nu=\nu_B=k^2=0} \frac{(p_1+p_2)_\alpha}{2M_N} \\
& \qquad + i\gamma_\alpha\bar{B}^{\pi N(-)}\bigg|_{\nu=\nu_B=k^2=0} \bigg] \tfrac{1}{2}[\tau^n,\tau^j] \bigg\} u(p_1).
\end{aligned}
\tag{38}
$$

[11] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. **106**, 1337 (1957). Note that, according to Eq. (34), $-k$ is the ingoing pion four-momentum.

Other derivative terms vanish at $\nu=0$ because of the well-known[11]) crossing properties of $A^{\pi N}$ and $B^{\pi N}$,

$$
\begin{aligned}
A^{\pi N(\pm)}(-\nu,\cdots) &= \pm A^{\pi N(\pm)}(\nu,\cdots), \\
B^{\pi N(\pm)}(-\nu,\cdots) &= \mp B^{\pi N(\pm)}(\nu,\cdots).
\end{aligned}
\tag{39}
$$

Since $-k^2$ is the (mass)^2 of the initial pion, Eq. (38) involves the pion-nucleon scattering amplitude extrapolated slightly off mass shell. Note that Eq. (36) is just the consistency condition on πN scattering,[12]

$$
\bar{A}^{\pi N(+)}\bigg|_{\nu=\nu_B=k^2=0} = \frac{g_r g_r(0)}{M_N}. \tag{40}
$$

Equations (35), (37), and (38) give the two leading terms in an expansion of $M_\alpha{}^{jn}$ in powers of k,

$$
M_\alpha{}^{jn} = M_\alpha{}^{jn\ \mathrm{ext}'} + \Delta M_\alpha{}^{jn} + O(k). \tag{41}
$$

Alternatively, we can use the analog of Eq. (30) to find the leading term in an expansion in powers of q (the soft pion limit). In this case, one would take T^P in Eq. (30) to be the axial-vector vertex. There will be an additional term in Eq. (30) arising from the equal-time commutator of the two axial-vector currents involved. Assuming the commutation relations postulated by Gell-Mann,[13] we find[14]

$$
M_\alpha{}^{jn} = M_\alpha{}^{jn\ \mathrm{ext}'} + \Delta M_\alpha{}^{jn\prime} + O(q), \tag{42}
$$

with

$$
\begin{aligned}
\Delta M_\alpha{}^{jn\prime} = i\frac{g_r(0)}{2M_N}\bar{u}(p_2)\bigg\{ & \frac{\mu^V}{g_A}\frac{(p_1+p_2)_\alpha}{2M_N} \\
& + i\gamma_\alpha\bigg[g_A - \frac{1}{g_A} - \frac{\mu^V}{g_A} \bigg] \bigg\} \tfrac{1}{2}[\tau^n,\tau^j] u(p_1), \\
& \qquad \mu^V = 3.70.
\end{aligned}
\tag{43}
$$

Clearly, at the point $q=k=0$ we must have $\Delta M_\alpha{}^{jn} = \Delta M_\alpha{}^{jn\prime}$. At this point $p_1=p_2$ and thus $i\gamma_\alpha$ and $(p_1+p_2)_\alpha/(2M_N)$ are equal between spinors. Hence, consistency between Eq. (42) and Eq. (41) demands

$$
1 - \frac{1}{g_A{}^2} = -\frac{2M^2{}_N}{g_r(0)^2}\bigg[\frac{\partial\bar{A}^{\pi N(-)}}{\partial\nu} + B^{\pi N(-)} \bigg]\bigg|_{\nu=\nu_B=k^2=q^2=0} \tag{44}
$$

which is the sum rule for the axial-vector coupling constant.[15]

[12] S. L. Adler, Phys. Rev. **137**, B1022 (1965).

[13] M. Gell-Mann, Physics **1**, 63 (1964).

[14] Y. Nambu and E. Shrauner, Phys. Rev. **128**, 862 (1962); S. L. Adler (to be published); G. Furlan, R. Jengo, and E. Remiddi, Nuovo Cimento **44**, 427 (1966). The diligent reader will actually find that in Eq. (42), and also in Eq. (45), we have dropped certain terms proportional to k_α which are not singular at $k^2=-m_\pi{}^2$. These terms are, of course, determined by our procedure, but they are numerically insignificant in weak pion production because k_α, contracted with the lepton current, becomes proportional to the lepton mass. We have also in Eq. (45) neglected a very small extra term, proportional to δ^{nj}, which appears in Eq. (41) when the pion four-momentum q is taken off mass shell [see W. I. Weisberger, Phys. Rev. **143**, 1302 (1966), Eq. (II.11a)].

[15] W. I. Weisberger, Phys. Rev. Letters **14**, 1047 (1965); S. L. Adler, Phys. Rev. Letters **14**, 1051 (1965).

Comparing Eqs. (43) and (38), we may determine the terms linear in either q or k. Our final result is then

$$M_\alpha{}^{jn} = M_\alpha{}^{jn\,\mathrm{ext}'} + \Delta M_\alpha{}^{jn''} + O(qk, q^2, k^2), \qquad (45)$$

with

$$\Delta M_\alpha{}^{jn''} = i\frac{M_N g_A}{g_r(0)}\bar{u}(p_2)\left\{\left[\frac{\partial \bar{A}^{\pi N(+)}}{\partial \nu_B}\bigg|_{\nu=\nu_B=k^2=q^2=0}\frac{q_\alpha}{2M_N}\right]\delta^{nj} + \left[\frac{g_r(0)^2}{2M^2{}_N}\left(1-\frac{1}{g_A{}^2}\right)\frac{(p_1+p_2)_\alpha}{2M_N} - i\frac{\sigma_{\alpha\beta}q_\beta}{2M_N}\bar{B}^{\pi N(-)}\bigg|_{\nu=\nu_B=k^2=q^2=0} + \frac{g_r(0)^2}{2M^2{}_N}\frac{i\sigma_{\alpha\beta}k_\beta}{2M_N} \times\left(1-\frac{1}{g_A{}^2}-\frac{\mu^V}{g_A{}^2}\right)\right]\tfrac{1}{2}[\tau^n,\tau^j]\right\}u(p_1). \qquad (46)$$

Unfortunately, it is doubtful if Eq. (46) will be of practical use, since there is a strong final-state interaction leading to the (3,3) resonance, which is located only one pion mass away from threshold in energy. This makes it unlikely that k and q will be good expansion parameters. However, we will use the same method of comparing expansions in q and k in dealing with radiative μ capture, where the final-state interaction is negligible and so the expansion may be physically interesting.

2. K_{e4} Decay

Here we consider the process

$$K^+(k^+) \to \pi^+(p^+) + \pi^-(p^-) + \bar{e}(k_e) + \nu(k_\nu). \qquad (47)$$

Again the four-momentum of each particle is indicated in parentheses. Let the four-momentum carried away by the lepton pair be k^-,

$$k_e + k_\nu = k^-. \qquad (48)$$

The most general form of the axial-vector contribution to the decay matrix element is

$$M_\alpha = (2k_0{}^+ 2p_0{}^+ 2p_0{}^-)^{1/2}\,{}_{\mathrm{out}}\langle \pi^+\pi^- | J_\alpha{}^{A,\,\Delta S=-1} | K^+\rangle = \frac{1}{m_K}[F_1(p^+ + p^-)_\alpha + F_2(p^+ - p^-)_\alpha + F_3 k_\alpha{}^-]. \qquad (49)$$

The form factors F are functions of the arguments $x = (p^+ + p^-)\cdot k^-$, $y = (p^+ - p^-)\cdot k^-$, and $(k^-)^2$. We define the matrix element for $\pi^+\pi^- \to K^+K^-$ by writing

$$(2k_0{}^+ 2p_0{}^+ 2p_0{}^-)^{1/2}\,{}_{\mathrm{out}}\langle \pi^+\pi^- | J_K | K^+\rangle_{\mathrm{in}} = iT_{\pi K}[x, y, (k^-)^2]. \qquad (50)$$

Then if we assume PCAC in the strangeness-changing case,[16]

$$\partial_\alpha J_\alpha{}^{A,\,\Delta S=-1} = C_K m_K{}^2 \phi_K, \qquad (51)$$

we find that

$$\frac{1}{m_K}F_1\bigg|_{x=y=(k^-)^2=0} = C_K\frac{\partial}{\partial x}T_{\pi K}[x, y, (k^-)^2]\bigg|_{x=y=(k^-)^2=0},$$
$$\frac{1}{m_K}F_2\bigg|_{x=y=(k^-)^2=0} = C_K\frac{\partial}{\partial y}T_{\pi K}[x, y, (k^-)^2]\bigg|_{x=y=(k^-)^2=0}. \qquad (52)$$

Hence, the K_{e4} decay amplitudes at a point on the boundary of the Dalitz plot are related to the πK amplitude, with one K meson off mass shell. In terms of the conventional Mandelstam variables, the point $x = y = (k^-)^2 = 0$ is

$$s = (p^+ + p^-)^2 = -m_K{}^2,$$
$$t = (k^- + p^+)^2 = -m_\pi{}^2, \qquad (53)$$
$$u = (k^- - p^-)^2 = -m_\pi{}^2.$$

3. Radiative μ Capture

In this subsection we discuss the process of radiative μ capture by a proton. This is an example of the situation, discussed briefly at the end of Sec. II, in which more than one current is acting. Consider then

$$\mu^-(k_\mu) + p(p_1) \to \nu(k_\nu) + \gamma(k) + n(p_2), \qquad (54)$$

and let

$$q = k_\mu - k_\nu \qquad (55)$$

be the lepton four-momentum transfer. The matrix element for this process is given by

$$T = \frac{G}{\sqrt{2}}\left\{-\langle n|J_\alpha{}^W|p\rangle \bar{u}_\nu\gamma_\alpha(1+\gamma_5) \times\frac{1}{i\gamma\cdot(k_\mu - k)+m_\mu} ie\gamma_\lambda\epsilon_\lambda{}^* u_\mu \frac{1}{(2k_0)^{1/2}} + \langle n\gamma|J_\alpha{}^W|p\rangle\bar{u}_\nu\gamma_\alpha(1+\gamma_5)u_\mu\right\}, \qquad (56)$$

with ϵ_λ the polarization vector of the photon and G the Fermi constant. The two contributions to T correspond, respectively, to radiation by the muon (which is negatively charged) and to radiation by the hadrons. The matrix element $\langle n|J_\alpha{}^W|p\rangle$ is given by

$$\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\langle n|J_\alpha{}^W|p\rangle = i\bar{u}(p_2)[F_1{}^V((q-k)^2)\gamma_\alpha - F_2{}^V((q-k)^2)\sigma_{\alpha\beta}(q-k)_\beta + g_A((q-k)^2)\gamma_\alpha\gamma_5 - ih_A((q-k)^2)\gamma_5(q-k)_\alpha]u(p_1). \qquad (57)$$

---

[16] R. P. Feynman, in *Symmetries in Elementary Particle Physics* (Academic Press Inc., New York, 1965), p. 158. The constant C_K is given by $C_K = (M_N + M_\Lambda)g_A{}^{\Lambda N}/g^{KN\Lambda}(0)$, with $g_A{}^{\Lambda N}$ the Λ beta-decay coupling constant and $g^{KN\Lambda}$ the $KN\Lambda$ coupling. For applications of partial conservation of the strangeness-conserving axial-vector current to K_{e4} decays, see C. G. Callan and S. B. Treiman, Phys. Rev. Letters **16**, 153 (1966) and M. Suzuki, *ibid.* **16**, 212 (1966).

Here, $F_1{}^V(t)$ and $F_2{}^V(t)$ are the isovector Dirac and Pauli electromagnetic form factors $[F_1{}^V(0)=1, F_2{}^V(0)=\mu^V/(2M_N)]$, $g_A(t)$ is the axial-vector form factor, and $h_A(t)$ is the induced pseudoscalar form factor. Applying PCAC to the one-nucleon vertex of the axial-vector current, we find that $h_A(t)$ may be written in the form

$$h_A(t)=\frac{2M_N g_A[g_r/g_r(0)]}{t+m_\pi{}^2}+r(t)\,, \tag{58a}$$

$$r(t)=\frac{1}{t}\left[2M_N g_A(t)-2M_N g_A\frac{m_\pi{}^2 g_r(t)+tg_r}{g_r(0)(t+m_\pi{}^2)}\right], \tag{58b}$$
$$r(0)\approx 2M_N g_A{}'(0)\,,$$

which explicitly exhibits the one-pion pole part and the remainder $r(t)$.

We write $\langle n\gamma|J_\alpha{}^W|p\rangle$ in the following form:

$$(2k_0p_{10}p_{20}/M^2{}_N)^{1/2}\langle n\gamma|J_\alpha{}^W|p\rangle=e\epsilon_\lambda{}^*M_{\lambda\alpha}\,. \tag{59}$$

We wish to use our knowledge of the divergences of the vector and axial-vector currents to calculate $M_{\lambda\alpha}$, up to and including terms linear in q and in k. In order to do this, we have to know the quantities $k_\lambda M_{\lambda\alpha}$ and $q_\alpha M_{\lambda\alpha}$. The first of these may be determined by conservation of the electromagnetic current. When $\epsilon_\lambda{}^*$ is replaced by k_λ in Eq. (56), the resulting expression must vanish. This tells us that

$$k_\lambda M_{\lambda\alpha}=-(p_{10}p_{20}/M^2{}_N)^{1/2}\langle n|J_\alpha{}^W|p\rangle\,. \tag{60}$$

In order to calculate $q_\alpha M_{\lambda\alpha}$, we made use of our knowledge of the divergences of the vector and the axial-vector parts of the weak current,[17]

$$J_\alpha{}^W=J_\alpha{}^V+J_\alpha{}^A\,, \tag{61a}$$

$$\partial_\alpha J_\alpha{}^V=ieA_\alpha J_\alpha{}^V\,,$$
$$\partial_\alpha J_\alpha{}^A=ieA_\alpha J_\alpha{}^A+(\sqrt{2}M_N m_\pi{}^2 g_A/g_r(0))\phi_{\pi^+}\,, \tag{61b}$$

where A_α is the electromagnetic vector potential and ϕ_{π^+} is the field which annihilates a positive pion. Equations (61b) follow from the assumption of minimal electromagnetic coupling and from the divergence equations in the absence of electromagnetism. (The factor $\sqrt{2}$ in the axial-vector equation comes from the definitions of $J_\alpha{}^A$ and ϕ_{π^+}: $J_\alpha{}^A=J_\alpha{}^{A1}-iJ_\alpha{}^{A2}$ and $\phi_{\pi^+}=(\phi_\pi{}^1-i\phi_\pi{}^2)/\sqrt{2}$.) Using Eqs. (61b) to evaluate $\langle n\gamma|\partial_\alpha J_\alpha{}^W|p\rangle$, we find

$$\epsilon_\lambda{}^*q_\alpha M_{\lambda\alpha}=-\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\epsilon_\lambda{}^*\langle n|J_\lambda{}^W|p\rangle$$
$$+i\frac{\sqrt{2}M_N g_A}{g_r(0)}\frac{m_\pi{}^2}{q^2+m_\pi{}^2}\left(2k_0\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}$$
$$\times e^{-1}\langle n\gamma|J_{\pi^+}|p\rangle\,. \tag{62}$$

[17] S. L. Adler, Phys. Rev. 139, B1638 (1965).

From Eqs. (60) and (62), we can deduce the gauge condition satisfied by

$$\left(2k_0\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\langle n\gamma|J_{\pi^+}|p\rangle=e\epsilon_\lambda{}^*T_{\pi^+\lambda}\,. \tag{63}$$

Replacing $\epsilon_\lambda{}^*$ by k_λ in Eq. (62), and multiplying Eq. (60) by q_α, we get

$$k_\lambda T_{\pi^+\lambda}=-\frac{q^2+m_\pi{}^2}{(q-k)^2+m_\pi{}^2}\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\langle n|J_{\pi^+}|p\rangle$$
$$=-\frac{q^2+m_\pi{}^2}{(q-k)^2+m_\pi{}^2}\sqrt{2}\bar{u}(p_2)i\gamma_5 g_r((q-k)^2)u(p_1)\,. \tag{64}$$

When $q^2=-m_\pi{}^2$, Eq. (64) becomes $k_\lambda T_{\pi^+\lambda}=0$, the usual gauge condition for on-mass-shell pion photoproduction.

Before stating the results for radiative μ capture, we will discuss the significance of Eqs. (60) and (62). A more conventional way to proceed in calculating $k_\lambda M_{\lambda\alpha}$ and $q_\alpha M_{\lambda\alpha}$ would be to contract the photon in Eq. (59), giving

$$eM_{\lambda\alpha}=\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}i\int d^4x\, e^{-ik\cdot x}(-\Box_x)$$
$$\times\langle n|T[A_\lambda(x)J_\alpha{}^W(0)]|p\rangle$$
$$=\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}i\left[\int d^4x\, e^{-ik\cdot x}\right.$$
$$\left.\times e\langle n|T[J_\lambda{}^{EM}(x)J_\alpha{}^W(0)]|p\rangle+S_{\lambda\alpha}\right], \tag{65}$$

with

$$S_{\lambda\alpha}=\int d^4x\, e^{-ik\cdot x}\delta(x_0)$$
$$\times\langle n|[\partial A_\lambda(x)/\partial x_0,J_\alpha{}^W(0)]|p\rangle\,, \tag{66}$$

where we have assumed that A_λ and $J_\alpha{}^W$ commute at equal times. The equal time commutator term $S_{\lambda\alpha}$ in Eq. (65), sometimes called a "seagull" or "catastrophic" term, describes the coupling of the weak and electromagnetic currents at the same point (see Fig. 4). It is a reflection of the extent to which A_λ appears in $J_\alpha{}^W$. Calculating $k_\lambda M_{\lambda\alpha}$, we now get

$$k_\lambda M_{\lambda\alpha}=\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\int dx_0 e^{ik_0x_0}\delta(x_0)$$
$$\times\left\langle n\left|\left[\int d^3x\, e^{-ik\cdot x}J_0{}^{EM}(x),J_\alpha{}^W(0)\right]\right|p\right\rangle$$
$$+\left(\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}e^{-1}ik_\lambda S_{\lambda\alpha}\,. \tag{67}$$

FIG. 4. A "seagull" diagram.

The commutator of the currents is

$$
\begin{aligned}
\delta(x_0)[J_0^{\mathrm{EM}}(x),J_\alpha^W(0)] \\
=-\delta^{(4)}(x)J_\alpha^W(0)+[&\text{possible gradient terms} \\
&\text{proportional to } \partial_\alpha\delta^{(3)}(\mathbf{x})]. \qquad (68)
\end{aligned}
$$

The first term in Eq. (68) is the one conjectured by Gell-Mann[13]; the possible presence of the gradient terms was pointed out by Schwinger.[18] We see that Eq. (60) implies that the Schwinger terms *exactly cancel* the divergence of the "seagull" terms. This cancellation has been proved by Feynman in a Yang-Mills theory and has been conjectured by him to be a general result.[19] In other words, when calculating the *divergence* of quantities like $M_{\lambda\alpha}$, if one neglects *both* the "seagull" terms and the Schwinger terms, one gets the right result. Note that the "seagull" terms cannot be dropped when calculating the matrix element $M_{\lambda\alpha}$ itself.

In order to state our answer for radiative μ capture, we have to define the amplitudes for pion photoproduction with the pion off-mass-shell. This process is related by crossing symmetry to the matrix element $\langle n\gamma|J_{\pi^+}|p\rangle$ in Eq. (62). We write the photoproduction amplitude in the following form,[20]

$$
\begin{aligned}
&\left(2k_0\frac{p_{10}p_{20}}{M^2{}_N}\right)^{1/2}\langle N|J_\pi|N\gamma\rangle \\
&=e\psi_j^*\chi_2^*\bar{u}(p_2)\Bigg\{ig_r(q^2)\tau^j\gamma_5\frac{1}{i\gamma\cdot(p_2+q)+M_N} \\
&\quad\times\tfrac{1}{2}i\left[\gamma_\lambda(1+\tau^3)-\frac{\sigma_{\lambda\xi}k_\xi}{2M_N}(\mu^S+\mu^V\tau^3)\right] \\
&\quad+\tfrac{1}{2}i\left[\gamma_\lambda(1+\tau^3)-\frac{\sigma_{\lambda\xi}k_\xi}{2M_N}(\mu^S+\mu^V\tau^3)\right] \\
&\quad\times\frac{1}{i\gamma\cdot(p_1-q)+M_N}ig_r(q^2)\tau^j\gamma_5 \\
&\quad+ig_r(-m_\pi^2)[\tau^j,\tau^3]\gamma_5\frac{\frac{1}{2}(2q-k)_\lambda}{(q-k)^2+m_\pi^2} \\
&\quad+\sum_{s=1}^{4}O_{s\lambda}[\bar{V}_s^{(+)}\tfrac{1}{2}\delta^{j3}+\bar{V}_s^{(-)}\tfrac{1}{4}[\tau^j,\tau^3]+\bar{V}_s^{(0)}\tfrac{1}{2}\tau^j] \\
&\quad-i\gamma_5[\tau^j,\tau^3]q_\lambda\left(1+\frac{q^2}{m_\pi^2}\right)\left[g_r'(0)+\frac{g_r(-m_\pi^2)-g_r(0)}{m_\pi^2}\right] \\
&\quad+\left(1+\frac{q^2}{m_\pi^2}\right)O(q^2)\Bigg\}u(p)\chi_1\epsilon_\lambda, \qquad (69)
\end{aligned}
$$

where ψ_j is the isospin wave function of the pion, χ_1 and χ_2 are the nucleon isospinors, k is the ingoing photon four-momentum, and q is the outgoing pion four-momentum. The isoscalar nucleon anomalous magnetic moment has been denoted by $\mu^S[2M_NF_2^S(0)=\mu^S=-0.12]$. The four-vectors $O_{s\lambda}$, which satisfy $k_\lambda O_{s\lambda}=0$, are given by

$$
\begin{aligned}
O_{1\lambda}&=\tfrac{1}{2}i\gamma_5(\gamma_\lambda\gamma\cdot k-\gamma\cdot k\gamma_\lambda), & \eta_1&=1 \\
O_{2\lambda}&=i\gamma_5[(p_1+p_2)_\lambda q\cdot k-(p_1+p_2)\cdot kq_\lambda], & \eta_2&=1 \\
O_{3\lambda}&=\gamma_5(\gamma_\lambda q\cdot k-\gamma\cdot kq_\lambda), & \eta_3&=-1 \qquad (70) \\
O_{4\lambda}&=\gamma_5[\gamma_\lambda(p_1+p_2)\cdot k-\gamma\cdot k(p_1+p_2)_\lambda] \\
&\qquad -iM_N\gamma_5(\gamma_\lambda\gamma\cdot k-\gamma\cdot k\gamma_\lambda). & \eta_4&=1
\end{aligned}
$$

The amplitudes $\bar{V}_s$ are functions of the invariants q^2, k^2, ν, and ν_B, with

$$\nu=-k\cdot(p_1+p_2)/2M_N,\quad \nu_B=q\cdot k/2M_N. \qquad (71)$$

The bar on top of the V_s is a reminder that the Born term has been separated off. The numbers η_s specify the crossing properties[20] of the amplitudes $\bar{V}_s$,

$$\bar{V}_s^{(\pm,0)}(-\nu,\cdots)=\eta_s(\pm1,1)\bar{V}_s^{(\pm,0)}(\nu,\cdots). \qquad (72)$$

The terms explicitly proportional to $(1+q^2/m_\pi^2)$ in Eq. (69) are necessary to satisfy Eq. (64), the gauge-invariance requirement when the pion is off-mass-shell. Since

$$g_r'(0)+\frac{g_r(-m_\pi^2)-g_r(0)}{m_\pi^2}\approx\frac{m_\pi^2}{2}g_r''(0), \qquad (73)$$

the gauge-invariance term is numerically very small. The matrix element $\langle\gamma n|J_{\pi^+}|p\rangle$, which is the one needed in Eq. (62), is obtained from Eq. (69) by the replacements

$$
\begin{aligned}
\psi_j^*\rightarrow\psi_j&=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\\0\end{pmatrix}_j, \\
\epsilon_\lambda&\rightarrow\epsilon_\lambda^*, \qquad (74)\\
q&\rightarrow -q, \\
k&\rightarrow -k.
\end{aligned}
$$

Since the final nucleon is a neutron and the initial one is a proton, we have

$$\chi_2=\begin{pmatrix}0\\1\end{pmatrix},\quad \chi_1=\begin{pmatrix}1\\0\end{pmatrix}. \qquad (75)$$

We can now state the result for radiative μ capture:

$$
\begin{aligned}
M_{\lambda\alpha}=M_{\lambda\alpha}{}^N+M_{\lambda\alpha}{}^{RPD}+M_{\lambda\alpha}{}^{PPP} \\
+M_{\lambda\alpha}{}^R+O\left(\frac{q^2}{m_R^2},\frac{qk}{m_R^2}\right). \qquad (76)
\end{aligned}
$$

The mass m_R, which characterizes the terms neglected in our calculation, will typically be several pion masses

---

[18] J. Schwinger, Phys. Rev. Letters 3, 296 (1959) and Phys. Rev. 130, 406 (1963).

[19] R. P. Feynman (private communication).

[20] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. 106, 1345 (1957). The amplitudes $(V_1,V_2,V_3,V_4)^{(\pm 0)}$, as defined in Eq. (69), are respectively *double* the corresponding amplitudes $(A,B,C,D)^{(\pm 0)}$ of CGLN. [The isospin matrix elements in Eq. (69) are one-half those of CGLN.]

or greater in magnitude, since we have explicitly included all pion propagator terms.[21] In Eq. (76), we have

$$M_{\lambda\alpha}{}^{N}=\bar{u}(p_2)\Big\{[ig_A(q^2)\gamma_\alpha\gamma_5+h_A(q^2)\gamma_5q_\alpha$$
$$+iF_1{}^V(q^2)\gamma_\alpha-iF_2{}^V(q^2)\sigma_{\alpha\beta}q_\beta]$$
$$\times\frac{1}{i\gamma\cdot(p_2-q)+M_N}i\left[\gamma_\lambda+\mu^p\frac{\sigma_{\lambda\xi}k_\xi}{2M_N}\right]$$
$$+i\left[\mu^n\frac{\sigma_{\lambda\xi}k_\xi}{2M_N}\right]\frac{1}{i\gamma\cdot(p_1+q)+M_N}$$
$$\times[ig_A(q^2)\gamma_\alpha\gamma_5+h_A(q^2)\gamma_5q_\alpha$$
$$+iF_1{}^V(q^2)\gamma_\alpha-iF_2{}^V(q^2)\sigma_{\alpha\beta}q_\beta]\Big\}u(p_1)\,,\quad(77)$$

$$M_{\lambda\alpha}{}^{RPD}=\frac{2M_Ng_A}{g_r(0)}g_r\frac{\bar{u}(p_2)\gamma_5u(p_1)}{(q-k)^2+m_\pi{}^2}$$
$$\times\left[\frac{-2q_\lambda q_\alpha}{q^2+m_\pi{}^2}+\delta_{\lambda\alpha}+(q_\lambda k_\alpha-q\cdot k\delta_{\lambda\alpha})S\right],\quad(78)$$

$$M_{\lambda\alpha}{}^{PPP}=-ig_A\frac{q_\alpha}{q^2+m_\pi{}^2}\bar{u}(p_2)$$
$$\times\left[\gamma_5\sigma_{\lambda\xi}k_\xi\frac{\mu^S}{2M_N}+\frac{M_N}{g_r(0)}q_\beta O_{\lambda\beta}\right]u(p_1)\,,\quad(79)$$

$$M_{\lambda\alpha}{}^{R}=i\bar{u}(p_2)\{-\sigma_{\alpha\lambda}(\mu^V/2M_N)$$
$$+2g_A{}'(0)(q_\lambda\gamma_\alpha+k_\alpha\gamma_\lambda-\delta_{\lambda\alpha}\gamma\cdot k)\gamma_5$$
$$-ir(0)\delta_{\lambda\alpha}\gamma_5+2F_1{}^{V\prime}(0)(q_\lambda\gamma_\alpha+k_\alpha\gamma_\lambda)$$
$$-2F_2{}^{V\prime}(0)(q_\lambda\sigma_{\alpha\beta}q_\beta-k_\alpha\sigma_{\lambda\xi}k_\xi)$$
$$+[M_Ng_A/g_r(0)]O_{\lambda\alpha}\}u(p_1)\,,\quad(80)$$

with

$$\mu^p=1.79\,,\quad\mu^n=-1.91\,,\qquad(81)$$

and with

$$O_{\lambda\alpha}=\gamma_5\sigma_{\lambda\xi}k_\xi\left[\frac{\partial\bar{V}_1{}^{(0)}}{\partial\nu_B}\bigg|_0\frac{k_\alpha}{2M_N}+\frac{\partial\bar{V}_1{}^{(-)}}{\partial\nu}\bigg|_0\frac{(p_1+p_2)_\alpha}{2M_N}\right]$$
$$+i\gamma_5[(p_1+p_2)_\lambda k_\alpha-(p_1+p_2)\cdot k\sigma_{\lambda\alpha}]\bar{V}_2{}^{(0)}|_0$$
$$+\gamma_5[\gamma_\lambda k_\alpha-\gamma\cdot k\delta_{\lambda\alpha}]\bar{V}_3{}^{(-)}|_0$$
$$+k_\sigma\epsilon_{\lambda\alpha\sigma\mu}\gamma_\mu\bar{V}_4{}^{(0)}|_0\,.\quad(82)$$

In Eq. (82), $|_0$ means evaluation of the $\bar{V}_j$ at the point $\nu=\nu_B=q^2=k^2=0$.

Let us now discuss the various terms in Eq. (76). The nucleon Born term $M_{\lambda\alpha}{}^N$ corresponds to the diagrams of Figs. 5(a)-(d). The term $M_{\lambda\alpha}{}^{RPD}$ describes radiative

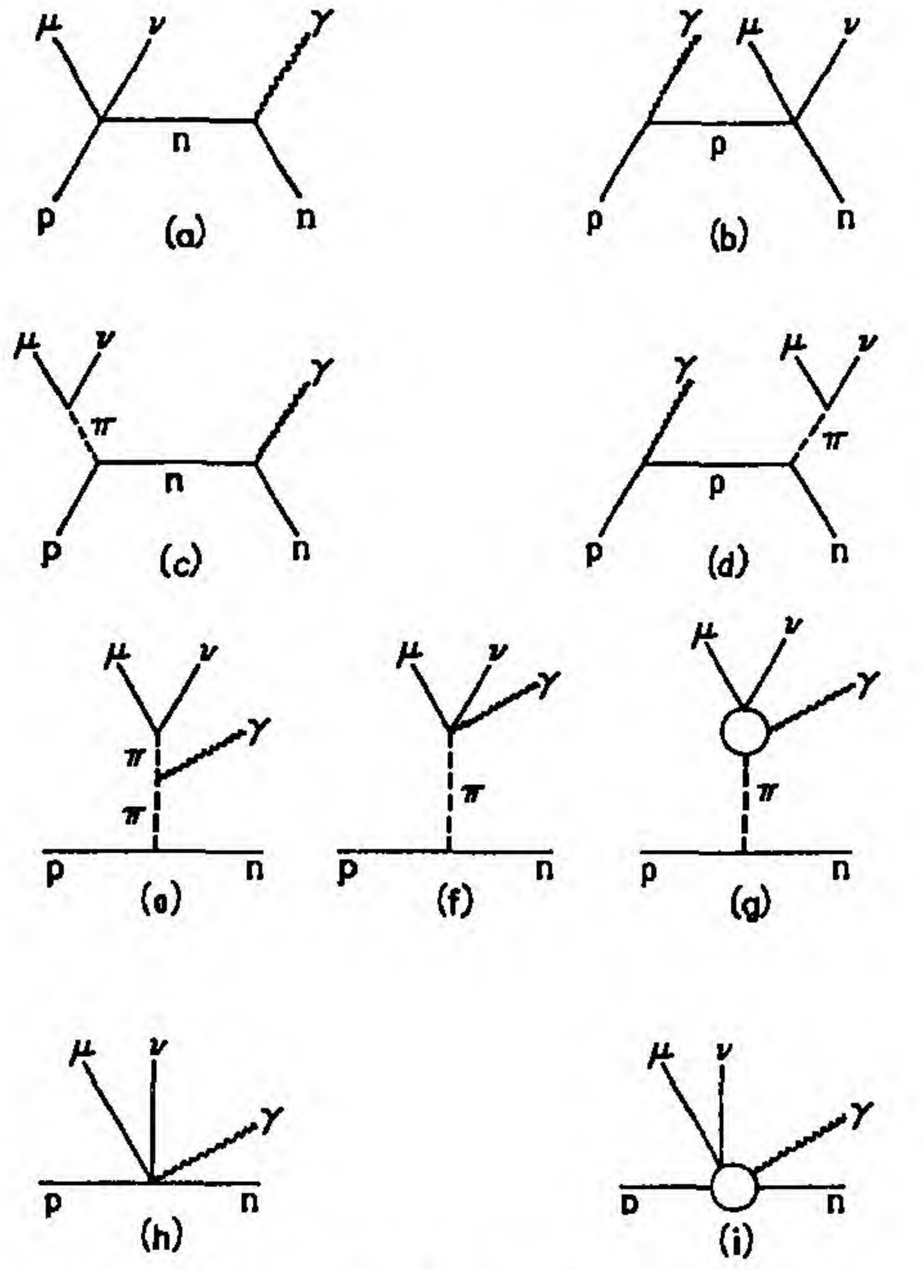


FIG. 5. Contributions to radiative μ capture.

virtual pion decay. The $q_\lambda q_\alpha$, $\delta_{\lambda\alpha}$, and S terms correspond, respectively, to Figs. 5(e)–5(g). The nontrivial structure term S *cannot* be determined by the procedure of this paper, because it is of order qk compared with the $\delta_{\lambda\alpha}$ term. The term $M_{\lambda\alpha}{}^{PPP}$ describes the structure part of virtual pion photoproduction. The Born part of virtual photoproduction has already been included in $M_{\lambda\alpha}{}^N$ and $M_{\lambda\alpha}{}^{RPD}$ [see Figs. 5(c)–5(e)]. In writing $M_{\lambda\alpha}{}^{PPP}$, we have eliminated $\bar{V}_1{}^{(0)}|_0$ by using Eq. (30), which implies

$$\bar{V}_1{}^{(0)}|_0=\frac{g_r(0)}{M_N}F_2{}^S(0)=\frac{g_r(0)\mu^S}{2M^2{}_N}\,.\qquad(83)$$

Equation (83) is one of the photoproduction sum rules derived by Fubini, Furlan, and Rossetti.[22]

The remainder term $M_{\lambda\alpha}{}^R$ is necessary to satisfy the divergence equations, Eq. (60) and Eq. (62). The first term, proportional to $\mu^V/2M_N$, has been included in previous calculations. It corresponds to the "seagull" diagram of Fig. 5(h). The remaining terms, linear in q or k, are new. They are represented diagrammatically by Fig. 5(i). We thus see that our procedure has allowed us to determine the leading nontrivial structure effects in radiative μ capture.

ACKNOWLEDGMENTS

We wish to thank F. J. Gilman and R. P Feynman for helpful discussions.

[21] The terms of order q^2 are determined by our procedure, but we have omitted them in writing the answer because they are as small numerically as the undetermined terms of order qk.

[22] S. Fubini, G. Furlan, and C. Rossetti, Nuovo Cimento **40**, 1171 (1965).

Low-Energy Theorem for the Weak Axial-Vector Vertex, S. L. ADLER AND Y. DOTHAN [Phys. Rev. **151**, 1267 (1966)]. In Eq. (80) for $M_{\lambda\alpha}{}^R$, the tensor multiplying $F_1{}^{V\prime}(0)$ should be $(q_\lambda\gamma_\alpha+k_\alpha\gamma_\lambda-\delta_{\lambda\alpha}\gamma\cdot k)$. In Eq. (82) for $O_{\lambda\alpha}$, the tensor multiplying $\bar{V}_2{}^{(0)}|_0$ should be $[(p_1+p_2)_\lambda k_\alpha-(p_1+p_2)\cdot k\delta_{\lambda\alpha}]$. Throughout Sec. III, $M^2{}_N$ should be read as $M_N{}^2$. We wish to thank Dr. J. Yellin for helpful discussions.

PHYSICAL REVIEW VOLUME 152, NUMBER 4 23 DECEMBER 1966

Partially Conserved Axial-Vector Current Restrictions on Pion Photoproduction and Electroproduction Amplitudes

STEPHEN L. ADLER*
Lyman Laboratory, Harvard University, Cambridge, Massachusetts

AND

FREDERICK J. GILMAN†
California Institute of Technology, Pasadena, California

(Received 4 August 1966)

We discuss numerically the restrictions imposed by the partially conserved axial-vector current (PCAC) on the pion photoproduction amplitude $V_1^{(+)}(0)$ and on the pion electroproduction amplitude $V_6^{(-)}(0)$. We find that the magnetic-dipole dominance and the narrow-resonance approximations are unreliable. The nonresonant s waves make an important contribution to $V_1^{(+)}(0)$, and we find that the PCAC prediction for this amplitude is reasonably well satisfied. The electric and longitudinal multipoles appear to make a much bigger contribution to $V_6^{(-)}(0)$ than does the magnetic dipole M_{1+}, which is strongly suppressed by the kinematics.

I. INTRODUCTION AND CONCLUSIONS

As has been much emphasized recently,[1] the partially conserved axial-vector current (PCAC) hypothesis, supplemented by current commutation relations, relates any weak or electromagnetic process in which a zero four-momentum pion is emitted to the same process in the absence of the pion. In particular, when applied to pion electroproduction, PCAC implies the relations[2]

$$(g_r(0)/M_N)F_2{}^V(k^2) = V_1{}^{(+)}(\nu=\nu_B=(M_\pi{}')^2=0,\,k^2)\,, \quad (1a)$$

$$(g_r(0)/M_N)F_2{}^S(k^2) = V_1{}^{(0)}(\nu=\nu_B=(M_\pi{}')^2=0,\,k^2)\,, \quad (1b)$$

$$\frac{g_r(0)}{M_N}\left[\frac{g_A(k^2)}{g_A(0)}-F_1{}^V(k^2)\right](k^2)^{-1} = V_6{}^{(-)}(\nu=\nu_B=(M_\pi{}')^2=0,\,k^2)\,. \quad (1c)$$

Here $F_1{}^V(k^2)$ is the isovector nucleon Dirac form factor; $F_2{}^V(k^2)$ and $F_2{}^S(k^2)$ are, respectively, the isovector and isoscalar nucleon Pauli form factors; $g_A(k^2)$ is the nucleon axial-vector form factor $[g_A(0)=1.18]$; and $g_r(0)$ is the pion-off-mass-shell pion-nucleon coupling constant $[g_r\equiv g_r(-M_\pi{}^2),\ g_r{}^2/4\pi\approx 14.6]$. The pion photoproduction amplitudes $V_1{}^{(+,0)}$ and the pion electroproduction amplitude $V_6{}^{(-)}$ will be specified more precisely below. When $k^2=0$, Eqs. (1a) and (1b) become the photoproduction relations of Fubini, Furlan, and Rossetti[3]; and Eq. (1c) becomes a relation between the axial-vector and charge radii of the nucleon.

The main purpose of this paper is to give a careful numerical analysis of Eqs. (1a) and (1c) at $k^2=0$. In the dispersion integrals for $V_1{}^{(+)}$ and $V_6{}^{(-)}$ we keep only the multipoles which resonate around the $N^*(1238)$ and the $N^{**}(1520)$, and the nonresonant s waves. As a preliminary, in Sec. II we state the needed kinematics and briefly derive Eqs. (1). In Sec. III we give the numerical discussion, using the photoproduction analyses of Schmidt and Höhler[4] and of Walker[5] in the region of the first two pion-nucleon resonances.

We reach the following conclusions:

1. The magnetic-dipole (M_{1+}) contribution to $V_1{}^{(+)}(0)$ from the neighborhood of the $N^*(1238)$ equals only about 0.75 times the left-hand side of Eq. (1a). Estimates based on the narrow-resonance approximation indicate a larger M_{1+} contribution, but we find that the narrow-resonance approximation for the $N^*(1238)$ overestimates integrals over the resonance by about 60%. When the resonant E_{1+}, M_{2-}, and E_{2-} multipoles are included, the value of $V_1{}^{(+)}(0)$ is reduced to about 0.6 times the left-hand side of Eq. (1a). However, the nonresonant s waves make a large contribution to the integral,[6] making the total integral for $V_1{}^{(+)}(0)$ equal to about 0.85 of the value predicted by PCAC.

2. The dispersion integral for $V_6{}^{(-)}(0)$ is *not* magnetic-dipole-dominated, because the M_{1+} contribution is kinematically suppressed. For instance, the multipole E_{1+} (electric quadrupole) in the $N^*(1238)$ region makes a contribution three times as big as the multipole M_{1+} to $V_6{}^{(-)}(0)$, even though the E_{1+} multipole is much

* Junior Fellow, Society of Fellows.
† National Science Foundation Postdoctoral Fellow, 1965–66.
[1] Y. Nambu and D. Lurié, Phys. Rev. **125**, 1429 (1962); Y. Nambu and E. Shrauner, *ibid.* **128**, 862 (1962); S. L. Adler, *ibid.* **139**, B1638 (1965); M. Suzuki, Phys. Rev. Letters **15**, 986 (1965); C. G. Callan and S. B. Treiman, *ibid.* **16**, 153 (1966).
[2] These relations are contained implicitly in the weak pion production results of Nambu and Shrauner (Ref. 1). The covariant forms have been derived by a number of authors: S. L. Adler, in Proceedings of the International Conference on Weak Interactions, Argonne National Laboratory, 1965, p. 291 (unpublished); Riazuddin and B. W. Lee, Phys. Rev. **146**, B1202 (1966); G. Furlan, R. Jengo, and E. Remiddi, Nuovo Cimento **44**, 427 (1966).
[3] S. Fubini, G. Furlan, and C. Rossetti, Nuovo Cimento **40**, 1171 (1965).
[4] W. Schmidt and G. Höhler, Ann. Phys. (N. Y.) **28**, 34 (1964); W. Schmidt, Z. Physik **182**, 76 (1964).
[5] R. L. Walker (private communication).
[6] The nonresonant s wave also makes an important contribution to the sum rule relating the isovector nucleon magnetic moment and charge radius to photoproduction cross sections—see F. J. Gilman and H. J. Schnitzer, Phys. Rev. **150**, 1362 (1966).

smaller than the M_{1+}. The value of $V_6^{(-)}(0)$ depends sensitively on the hard-to-measure longitudinal multipoles. Under the dubious assumption that the known proportionality of longitudinal and electric multipoles for zero photon momentum holds unchanged for large photon momenta as well, Eq. (1c) predicts an axial-vector form factor which falls off somewhat more slowly with k^2 than does $F_1^V(k^2)$.

The results of this paper should not be regarded as final, since the input multipole data may change as better analyses of photoproduction become available. What is definitely indicated, however, is that a comparison of Eqs. (1) with experiment must avoid unreliable narrow-resonance and M_{1+}-dominance approximations.

II. KINEMATICS AND DERIVATION OF PCAC RELATIONS

A. Kinematics

Let us consider the reaction

$$\gamma(k)+N(p_1)\rightarrow\pi(q)+N(p_2), \tag{2}$$

where the initial gamma may be real or virtual. The external particle masses are, respectively,

$$-k^2,\quad -p_1^2=M_N^2,\quad -q^2=(M_\pi{}^f)^2,\quad -p_2^2=M_N^2. \tag{3}$$

We define invariant-energy and momentum-transfer variables ν and ν_B by

$$\nu=-(p_1+p_2)\cdot k/(2M_N),\quad \nu_B=q\cdot k/(2M_N); \tag{4}$$

these are related to W, the invariant mass of the final pion-nucleon system, by

$$\nu-\nu_B=(W^2-M_N^2)/(2M_N). \tag{5}$$

All noninvariant quantities used in this paper refer to the reaction center-of-mass frame, in which $\mathbf{k}+\mathbf{p}_1=\mathbf{q}+\mathbf{p}_2=0$. We denote by y the cosine of the angle between the photon and pion directions:

$$y=\hat{q}\cdot\hat{k}, \tag{6}$$

and by $|\mathbf{k}|=(k_0^2+k^2)^{1/2}$ and $|\mathbf{q}|=(q_0^2-(M_\pi{}^f)^2)^{1/2}$ the photon and pion momenta. The photon and pion energies are given by

$$k_0=\frac{W^2-M_N^2-k^2}{2W},\quad q_0=\frac{W^2-M_N^2+(M_\pi{}^f)^2}{2W}. \tag{7}$$

The matrix element for the electroproduction reaction of Eq. (2) takes the form

$$m=e_r\epsilon_\lambda\,{}_{\text{out}}\langle\pi(q)N(p_2)|J_\lambda{}^{I3}+J_\lambda{}^Y|N(p_1)\rangle, \tag{8}$$

with e_r the electric charge, ϵ_λ the virtual photon polarization vector (which satisfies $k\cdot\epsilon=0$), and with $J_\lambda{}^{I3}$ and $J_\lambda{}^Y$, respectively, the third component of the isospin current and the hypercharge current. The isospin structure of the matrix element is given by

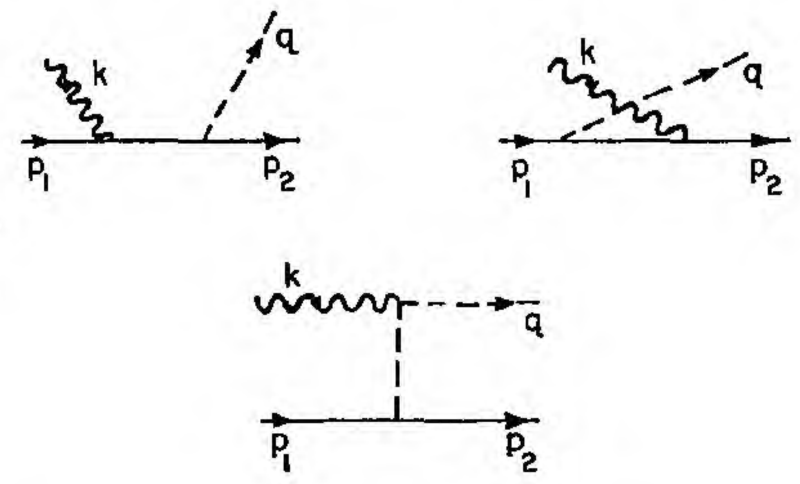


FIG. 1. Born approximation diagrams.

$$\begin{aligned}{}_{\text{out}}\langle\pi N|J_\lambda{}^{I3}|N\rangle&=a^{(+)}V_\lambda{}^{(+)}+a^{(-)}V_\lambda{}^{(-)},\\ {}_{\text{out}}\langle\pi N|J_\lambda{}^Y|N\rangle&=a^{(0)}V_\lambda{}^{(0)},\end{aligned} \tag{9}$$

with[7]

$$\begin{aligned}a^{(\pm)}&=\chi_f{}^{I*}\psi_c{}^*\tfrac{1}{4}(\tau_c\tau_3\pm\tau_3\tau_c)\chi_i{}^I,\\ a^{(0)}&=\chi_f{}^{I*}\psi_c{}^*\tfrac{1}{2}\tau_c\chi_i{}^I.\end{aligned} \tag{10}$$

In Eq. (10), ψ_c, $\chi_f{}^I$, and $\chi_i{}^I$ are, respectively, the isospinors of the final pion, the final nucleon, and the initial nucleon. The space-spin structure of the matrix element is given by

$$\epsilon_\lambda V_\lambda{}^{(\pm,0)}=\sum_{j=1}^{6}V_j{}^{(\pm,0)}(\nu,\nu_B,(M_\pi{}^f)^2,k^2)\times\bar{u}(p_2)O(V_j)u(p_1). \tag{11}$$

Defining $\{a,b\}=a\cdot\epsilon b\cdot k-a\cdot kb\cdot\epsilon$, we may take the $O(V_j)$ as

$$\begin{aligned}
O(V_1)&=\tfrac{1}{2}i\gamma_5\{\gamma,\gamma\}, & \eta_1{}^V&=1;\\
O(V_2)&=i\gamma_5\{p_1+p_2,\,q\}, & \eta_2{}^V&=1;\\
O(V_3)&=\gamma_5\{\gamma,q\}, & \eta_3{}^V&=-1;\\
O(V_4)&=\gamma_5\{\gamma,\,p_1+p_2\}-iM_N\gamma_5\{\gamma,\gamma\}, & \eta_4{}^V&=1;\\
O(V_5)&=i\gamma_5\{k,q\}, & \eta_5{}^V&=-1;\\
O(V_6)&=\gamma_5\{k,\gamma\}, & \eta_6{}^V&=-1.
\end{aligned} \tag{12}$$

The numbers $\eta_j{}^V$ specify the crossing properties of the invariant amplitudes:

$$V_j{}^{(\pm,0)}(\nu,\nu_B,(M_\pi{}^f)^2,k^2)=(\pm,+)\eta_j{}^VV_j{}^{(\pm,0)}(-\nu,\,\nu_B,\,(M_\pi{}^f)^2,\,k^2). \tag{13}$$

To make the normalization precise, we state the contribution of the Born approximation diagrams of Fig. 1 to the invariant amplitudes. [In the following equations we take the external pion to be physical

[7] Our notation follows that of a review article on pion electro- and weak production in preparation by one of the authors (S.L.A.). Our amplitudes are related to those of CGLN [G. F. Chew, F. E. Low, M. L. Goldberger, and Y. Nambu, Phys. Rev. **106**, 1345 (1957)] as follows:

covariant amplitudes—$[V_1,V_2,V_3,V_4]^{(\pm,0)}{}_{\text{this paper}}=2[A,B,C,D]^{(\pm,0)}{}_{\text{CGLN}}$,

center-of-mass amplitudes—$[\mathfrak{F}_j{}^V]^{(\pm,0)}{}_{\text{this paper}}=(8\pi W/M_N)[\mathfrak{F}_j{}^V]^{(\pm,0)}{}_{\text{CGLN}}$,

multipoles—$[M_{l+},\text{etc.}]^{(\pm,0)}{}_{\text{this paper}}=(8\pi W/M_N)[M_{l+},\text{etc.}]^{(\pm,0)}{}_{\text{CGLN}}$.

$(M_\pi^f=M_\pi)$; $F_\pi(k^2)$ is the pion charge form factor.]

$$V_1^{(\pm)B}=-\frac{g_rF_1^V(k^2)}{2M_N}\left(\frac{1}{\nu_B-\nu}\pm\frac{1}{\nu_B+\nu}\right),$$
$$V_2^{(\pm)B}=\frac{g_rF_1^V(k^2)}{4M_N^2\nu_B}\left(\frac{1}{\nu_B-\nu}\pm\frac{1}{\nu_B+\nu}\right),$$
$$V_3^{(\pm)B}=\frac{g_rF_2^V(k^2)}{2M_N}\left(\frac{1}{\nu_B-\nu}\mp\frac{1}{\nu_B+\nu}\right),\qquad(14)$$
$$V_4^{(\pm)B}=\frac{g_rF_2^V(k^2)}{2M_N}\left(\frac{1}{\nu_B-\nu}\pm\frac{1}{\nu_B+\nu}\right),$$
$$V_5^{(+)B}=0,\quad V_5^{(-)B}=\frac{-2g_r}{k^2}\left(\frac{F_1^V(k^2)}{2M_N\nu_B}-\frac{2F_\pi(k^2)}{4M_N\nu_B-k^2}\right),$$
$$V_6^{(\pm)B}=0.$$

While the consequences of PCAC are most simply expressed in terms of the invariant amplitudes V_j, pion photoproduction and electroproduction experiments are most easily analyzed in terms of the center-of-mass frame amplitudes $\mathfrak{F}_j^V$, defined by[7]

$$\epsilon_\lambda V_\lambda^{(\pm,0)}=\sum_{j=1}^{6}\mathfrak{F}_j^{V(\pm,0)}\chi_f^*\Sigma_j^V\chi_i.\qquad(15)$$

Here χ_f and χ_i are the nucleon Pauli spinors, and the Σ's are chosen as follows:

$$\Sigma_1^V=i(\sigma\cdot\varepsilon-\sigma\cdot\hat{k}\hat{k}\cdot\varepsilon),\quad \Sigma_4^V=i\sigma\cdot\hat{q}(\hat{q}\cdot\varepsilon-\hat{q}\cdot\hat{k}\hat{k}\cdot\varepsilon),$$
$$\Sigma_2^V=\sigma\cdot\hat{q}\sigma\cdot(\hat{k}\times\varepsilon),\quad \Sigma_5^V=-ik^2\sigma\cdot\hat{k}\hat{k}\cdot\varepsilon/k_0,\qquad(16)$$
$$\Sigma_3^V=i\sigma\cdot\hat{k}(\hat{q}\cdot\varepsilon-\hat{q}\cdot\hat{k}\hat{k}\cdot\varepsilon),\quad \Sigma_6^V=-ik^2\sigma\cdot\hat{q}\hat{k}\cdot\varepsilon/k_0.$$

The amplitudes $\mathfrak{F}_j^V$ have simple multipole expansions[7]:

$$\mathfrak{F}_1^V=\sum_{l=0}^{\infty}(lM_{l+}+E_{l+})P_{l+1}'(y)+\sum_{l=2}^{\infty}[(l+1)M_{l-}+E_{l-}]P_{l-1}'(y),$$
$$\mathfrak{F}_2^V=\sum_{l=1}^{\infty}[(l+1)M_{l+}+lM_{l-}]P_l'(y),$$
$$\mathfrak{F}_3^V=\sum_{l=1}^{\infty}(-M_{l+}+E_{l+})P_{l+1}''(y)+\sum_{l=3}^{\infty}(M_{l-}+E_{l-})P_{l-1}''(y),\qquad(17)$$
$$\mathfrak{F}_4^V=\sum_{l=2}^{\infty}(M_{l+}-M_{l-}-E_{l+}-E_{l-})P_l''(y),$$
$$k_0\mathfrak{F}_5^V=\sum_{l=0}^{\infty}(l+1)L_{l+}P_{l+1}'(y)-\sum_{l=2}^{\infty}lL_{l-}P_{l-1}'(y),$$
$$k_0\mathfrak{F}_6^V=\sum_{l=1}^{\infty}[lL_{l-}-(l+1)L_{l+}]P_l'(y).$$

The index $l\pm$ of the multipole specifies the orbital angular momentum (l) and the total angular momentum ($J=l\pm\frac{1}{2}$) of the final pion-nucleon system. It is straightforward, but tedious, to calculate the linear transformations connecting the amplitudes V_j and $\mathfrak{F}_j^V$.[8]

B. Derivation

The PCAC relations of Eq. (1) come from the identity

$$i\int d^4x\,e^{-iq\cdot x}\psi_c^*(-\Box_x+M_\pi^2)\langle N(p_2)|T[\partial_\sigma J_\sigma^{Ac}(x),(J_\lambda^{I3}(0)+J_\lambda^Y(0))]|N(p_1)\rangle\epsilon_\lambda$$
$$=-i\int d^4x\,e^{-iq\cdot x}\psi_c^*(-\Box_x+M_\pi^2)\delta(x_0)\langle N(p_2)|[J_0^{Ac}(x),J_\lambda^{I3}(0)+J_\lambda^Y(0)]|N(p_1)\rangle\epsilon_\lambda$$
$$-q_\sigma\int d^4x\,e^{-iq\cdot x}\psi_c^*(-\Box_x+M_\pi^2)\langle N(P_2)|T[J_\sigma^{Ac}(x),(J_\lambda^{I3}(0)+J_\lambda^Y(0))]|N(p_1)\rangle\epsilon_\lambda,\qquad(18)$$

which is obtained by integration by parts. Using the partially conserved axial-vector current hypothesis,[9]

$$\partial_\sigma J_\sigma^{Ac}(x)=\frac{M_NM_\pi^2g_A}{g_r(0)}\varphi_\pi^c(x),\qquad(19)$$

we see that the left-hand side of Eq. (18) is just

$$\frac{M_NM_\pi^2g_A}{g_r(0)}\sum_{j=1}^{6}\bar{u}(p_2)O(V_j)u(p_1)[a^{(+)}\bar{V}_j^{(+)}+a^{(-)}\bar{V}_j^{(-)}+a^{(0)}\bar{V}_j^{(0)}]+\text{Born terms},\qquad(20)$$

where $\bar{V}_j$ denotes the non-Born part of the amplitude V_j.

[8] See, for example, R. Blankenbecler, S. Gartenhaus, R. Huff, and Y. Nambu, Nuovo Cimento 17, 775 (1960); P. Dennery, Phys. Rev. 124, 2000 (1961).
[9] M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960); Y. Nambu, Phys. Rev. Letters 4, 380 (1960).

Let us evaluate the two terms on the right-hand side of Eq. (18) in the limit as $q\to 0$. The equal-time commutator term approaches

$$-iM_\pi^2\psi_c^*\langle N(p_2)|\left[\int d^3xJ_0^{Ac}(x),\, J_\lambda^{I3}(0)+J_\lambda^{Y}(0)\right]\Big|_{x_0=0}|N(p_1)\rangle\epsilon_\lambda. \tag{21}$$

Because of the integration over all space, possible gradient terms in the commutator do not contribute, and we find for this term

$$(M_\pi^2g_A(k^2)/k^2)a^{(-)}\bar{u}(p_2)O(V_6)u(p_1). \tag{22}$$

(To simplify the algebra we have dropped terms proportional to $k\cdot\epsilon=0$.) The term proportional to q_σ, in the limit as $q_\sigma\to 0$, can be evaluated by keeping only the one-nucleon-pole terms.[10] This gives

$$\begin{aligned}
&-M_\pi^2\psi_c^*\bar{u}(p_2)\Bigg\{ig_A\gamma\cdot q\gamma_5\frac{\tau_c}{2}\frac{\gamma\cdot p_2+iM_N}{-2p_2\cdot q}\tfrac{1}{2}i[\gamma_\lambda(F_1^V(k^2)\tau_3+F_1^S(k^2))-\sigma_{\lambda\mu}k_\mu(F_2^V(k^2)\tau_3+F_2^S(k^2))]\\
&\quad+\tfrac{1}{2}i[\gamma_\lambda(F_1^V(k^2)\tau_3+F_1^S(k^2))-\sigma_{\lambda\mu}k_\mu(F_2^V(k^2)\tau_3+F_2^S(k^2))]\frac{\gamma\cdot p_1+iM_N}{2p_1\cdot q}ig_A\gamma\cdot q\gamma_5\frac{\tau_c}{2}\Bigg\}u(p_1)\epsilon_\lambda\\
&=M_\pi^2g_A\Bigg\{-a^{(-)}\frac{F_1^V(k^2)}{k^2}\bar{u}(p_2)O(V_6)u(p_1)+(a^{(+)}F_2^V(k^2)+a^{(0)}F_2^S(k^2))\bar{u}(p_2)O(V_1)u(p_1)\\
&\quad-\frac{F_1^V(k^2)}{2}\left[a^{(+)}\left(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\right)+a^{(-)}\left(\frac{1}{\nu_B-\nu}-\frac{1}{\nu_B+\nu}\right)\right]\bar{u}(p_2)O(V_1)u(p_1)+\text{the other Born terms}\Bigg\}.
\end{aligned}\tag{23}$$

Comparing Eq. (20) with Eqs. (22) and (23), we get the relations

$$\begin{aligned}
&(g_r(0)/M_N)F_2^V(k^2)=\bar{V}_1^{(+)}(\nu=\nu_B=(M_\pi{}^f)^2=0,k^2),\\
&(g_r(0)/M_N)F_2^S(k^2)=\bar{V}_1^{(0)}(\nu=\nu_B=(M_\pi{}^f)^2=0,k^2),\\
&\frac{g_r(0)}{M_N}\left[\frac{g_A(k^2)}{g_A(0)}-F_1^V(k^2)\right](k^2)^{-1}=\bar{V}_6^{(-)}(\nu=\nu_B=(M_\pi{}^f)^2=0,k^2).
\end{aligned}\tag{24}$$

If we take $\nu=\nu_B=0$ to mean "first set $\nu_B=0$, *then* set $\nu=0$" the bars in Eq. (24) may be dropped, since the Born approximation to V_1 vanishes at $\nu_B=0$ (for all $\nu\neq0$). This completes the deviation of Eqs. (1).

III. NUMERICAL ANALYSIS

We now proceed to a numerical analysis of Eqs. (1a) and (1c) at $k^2=0$. Introducing the abbreviations $V_1^{(+)}(0)\equiv V_1^{(+)}(\nu=\nu_B=(M_\pi{}^f)^2=k^2=0)$, $V_6^{(-)}(0)\equiv V_6^{(-)}(\nu=\nu_B=(M_\pi{}^f)^2=k^2=0)$, we write the equations in the form

$$\frac{g_r}{M_N}F_2^V(0)=\frac{g_r}{g_r(0)}V_1^{(+)}(0),\quad \frac{g_r}{M_N}\left[\frac{g_A'(0)}{g_A(0)}-F_1^{V\prime}(0)\right]=\frac{g_r}{g_r(0)}V_6^{(-)}(0). \tag{25}$$

In order to calculate $V_1^{(+)}$ and $V_6^{(-)}$ from experimental photoproduction data, we assume that $V_1^{(+)}$ and $V_6^{(-)}$ both satisfy unsubtracted fixed-momentum-transfer dispersion relations in the energy variable ν[11]:

$$\begin{aligned}
V_1^{(+)}(\nu,\nu_B,(M_\pi{}^f)^2,k^2)=&\frac{-g_r(-(M_\pi{}^f)^2)F_1^V(k^2)}{2M_N}\left(\frac{1}{\nu_B-\nu}+\frac{1}{\nu_B+\nu}\right)\\
&+\frac{1}{\pi}\int_{\nu_B+M_\pi+M_\pi^2/(2M_N)}^{\infty}d\nu'\,\mathrm{Im}V_1^{(+)}(\nu',\nu_B,(M_\pi{}^f)^2,k^2)\left(\frac{1}{\nu'-\nu}+\frac{1}{\nu'+\nu}\right),
\end{aligned}\tag{26}$$

---

[10] See S. L. Adler, Ref. 1, where the rules for calculating the "pole insertions" are discussed. In Sec. II B we have ignored questions of gauge invariance. It is easily shown [S. Adler and Y. Dothan, Phys. Rev. **151**, 1267 (1966), and M. Nauenberg, Phys. Letters **22**, 201 (1966)] that when the final pion is off mass shell, the photoproduction or electroproduction amplitude is not divergenceless, but has a divergence proportional to $(q^2+M_\pi^2)g_r[(q-k)^2]/[(q-k)^2+M_\pi^2]$. In order to maintain the correct divergence, additional terms must be added to the Born approximation calculated from the diagrams of Fig. 1. However, these additional terms vanish when $q=k\cdot\epsilon=0$, and thus do not affect the results of this paper. See also S. Fubini, Y. Nambu, and A. Wataghin, Phys. Rev. **111**, 329 (1958).

[11] Validity of the unsubtracted dispersion relation for $V_1^{(+)}(0)$ is indicated by the Regge-pole analysis of photoproduction given by G. Zweig, Nuovo Cimento **32**, 689 (1964).

$$V_6^{(-)}(\nu,\nu_B,(M_\pi^f)^2,k^2)=\frac{1}{\pi}\int_{\nu_B+M_\pi+M_\pi^2/(2M_N)}^{\infty}d\nu'\,\mathrm{Im}V_6^{(-)}(\nu',\nu_B,(M_\pi^f)^2,k^2)\left(\frac{1}{\nu'-\nu}+\frac{1}{\nu'+\nu}\right),$$

which imply that

$$\frac{g_r}{g_r(0)}\left\{\begin{matrix}V_1^{(+)}(0)\\ V_6^{(-)}(0)\end{matrix}\right\}=\frac{2}{\pi}\int_{M_\pi+M_\pi^2/2M_N}^{\infty}\frac{d\nu'}{\nu'}\frac{g_r}{g_r(0)}\mathrm{Im}\left\{\begin{matrix}V_1^{(+)}(\nu',0,0,0)\\ V_6^{(-)}(\nu',0,0,0)\end{matrix}\right\}. \tag{27}$$

In order to calculate the integrand of Eq. (27), we make a multipole expansion, keeping only those multipoles which can at present be determined from experiment. These are: (i) the nonresonant s-wave multipoles E_{0+} and L_{0+}. The E_{0+} makes a large contribution to charged pion photoproduction; (ii) the multipoles $M_{1+}^{(3/2)}$, $E_{1+}^{(3/2)}$, and $L_{1+}^{(3/2)}$, which are important around the $I=\frac{3}{2}N^*(1238)$; (iii) the multipoles $M_{2-}^{(1/2)}$, $E_{2-}^{(1/2)}$, and $L_{2-}^{(1/2)}$ which are important around the $I=\frac{1}{2}N^{**}(1520)$.

Doing the necessary arithmetic, we find

$$\frac{g_r}{g_r(0)}V_1^{(+)}(\nu,0,0,0)=\frac{2M_N}{W^2-M_N^2}\frac{g_r}{g_r(0)}\left[\tfrac{1}{3}E_{0+}^{(1/2)}+\tfrac{2}{3}E_{0+}^{(3/2)}+\tfrac{2}{3}M_{1+}^{(3/2)}+2E_{1+}^{(3/2)}-M_{2-}^{(1/2)}+\tfrac{1}{3}E_{2-}^{(1/2)}\right]\Big|_{M_\pi^f=0}, \tag{28a}$$

$$\frac{g_r}{g_r(0)}V_6^{(-)}(\nu,0,0,0)=\frac{2M_N}{W^2-M_N^2}\frac{1}{3}\frac{g_r}{g_r(0)}\left[\frac{E_{0+}^{(1/2)}-E_{0+}^{(3/2)}}{W-M_N}-\frac{2W(L_{0+}^{(1/2)}-L_{0+}^{(3/2)})}{W^2-M_N^2}+\frac{M_{1+}^{(3/2)}}{W+M_N}\right.$$
$$\left.-\frac{3(3W+M_N)E_{1+}^{(3/2)}}{W^2-M_N^2}+\frac{8WL_{1+}^{(3/2)}}{W^2-M_N^2}+\frac{3M_{2-}^{(1/2)}}{W+M_N}+\frac{(M_N-5W)E_{2-}^{(1/2)}}{W^2-M_N^2}-\frac{8WL_{2-}^{(1/2)}}{W^2-M_N^2}\right]\Big|_{M_\pi^f=0}. \tag{28b}$$

The multipoles appearing in Eq. (28) are not actually the physical multipoles, since they refer to zero final pion mass ($M_\pi^f=0$). We relate them to the physical multipoles by the prescription

$$\left\{\begin{matrix}M\\E\\L\end{matrix}\right\}_{l\pm}^{(I)}\Bigg|_{M_\pi^f=0}=\frac{g_r(0)}{g_r}\left\{\begin{matrix}M\\E\\L\end{matrix}\right\}_{l\pm}^{(I)}\Bigg|_{M_\pi^f=M_\pi}\left(\frac{|\mathbf{q}|_{M_\pi^f=0}}{|\mathbf{q}|_{M_\pi^f=M_\pi}}\right)^l, \tag{29}$$

where the subscripts on $|\mathbf{q}|$ indicate that $|\mathbf{q}|$ is to be computed from W with $M_\pi^f=0$ or M_π, respectively. The prescription of Eq. (29) gives the unphysical multipoles the correct threshold behavior and, approximately, the correct nearby left-hand singularities.[12] Using Eq. (29) eliminates the obnoxious factor $g_r/g_r(0)$ in Eq. (28) and leaves us with simple integrals over the physically measurable multipoles.

From pion-photoproduction experiments, the electric and magnetic multipoles can be measured. However, the longitudinal multipoles can only be measured in pion electroproduction experiments; so far little data is available. Consequently, we will have to make a guess as to the magnitude of the longitudinal multipoles. When the photon momentum $|\mathbf{k}|$ approaches zero, the longitudinal and electric multipoles become proportional with known coefficients,[13]

$$L_{l+}/E_{l+}\to 1,\quad l\geq 0,$$
$$L_{l-}/E_{l-}\to -(l-1)/l,\quad l\geq 2. \tag{30}$$

[12] For a more detailed discussion see S. L. Adler, Phys. Rev. **140**, B736 (1965).

[13] J. D. Bjorken and J. D. Walecka, Ann. Phys. (N. Y.) 38, 35 (1966); Y. Dothan and R. P. Feynman (private communciation).

For want of a better estimate, we will assume that these proportionalities hold for nonzero $|\mathbf{k}|$ as well. In other words, we take

$$L_{0+}\approx E_{0+},\quad L_{1+}\approx E_{1+},\quad L_{2-}\approx -\tfrac{1}{2}E_{2-} \tag{31}$$

in the numerical work described below.

A. Narrow-Resonance Approximation

We begin by discussing the narrow-resonance approximation for the magnetic dipole ($M_{1+}^{(3/2)}$) contribution. It is convenient here to use the CGLN model[14] for $M_{1+}^{(3/2)}$, which, as Schmidt and Höhler[4] and Schmidt[4] have shown, is in good agreement with photoproduction experiments. According to this model

$$M_{1+}^{(3/2)}=\frac{4.70g_r}{3}\frac{W}{M_N}\frac{|\mathbf{q}||\mathbf{k}|}{M_N^2}\frac{\exp(i\delta_{3,3})\sin\delta_{3,3}}{\frac{4}{3}f^2|\mathbf{q}|^3/M_\pi^2}, \tag{32}$$

with $f^2=0.08$, $\delta_{3,3}$ the pion-nucleon scattering phase shift in the (3,3) partial wave, and $|\mathbf{q}|$ evaluated with

[14] G. F. Chew, F. E. Low, M. L. Goldberger, and Y. Nambu, Phys. Rev. **106**, 1345 (1957).

TABLE I. Parameters for multipoles.

Resonance	W_R (units of M_π)	$\|\mathbf{q}\|_R$ (units of M_π)	Γ_R (units of M_π)	Multipole $\mathfrak{M}$	A (units of M_π^{-1})
$N^*(1238)$	8.85	1.65	0.860	$M_{1+}^{(3/2)}$	+0.112
				$E_{1+}^{(3/2)}$	−0.0080
$N^{**}(1520)$	10.80	3.20	0.860	$M_{2-}^{(1/2)}$	+0.0155
				$E_{2-}^{(1/2)}$	+0.0628

$M_\pi{}^f=M_\pi$. Substituting Eq. (32) into Eq. (27), we find for the magnetic-dipole contribution to $V_1^{(+)}(0)$

$$V_1^{(+)}(0)|_{\text{magnetic dipole}}=\frac{g_r}{M_N}\frac{8}{9}\frac{4.70}{2M_N}I,$$
$$I=\frac{1}{\pi}\int_{M_N+M_\pi}^{\infty}dW\frac{\sin^2\delta_{3,3}}{[|\mathbf{q}|_{M_\pi{}^f=M_\pi}]^3\frac{4}{3}f^2/M_\pi{}^2}. \tag{33}$$

According to the narrow-resonance approximation,[15] $I=1$, giving

$$V_1^{(+)}(0)|_{\text{magnetic dipole (narrow resonance)}}\approx 0.62/M_\pi{}^2, \tag{34}$$

to be compared with the value predicted by PCAC [the left-hand side of Eq. (25)],

$$(g_r/M_N)F_2{}^V(0)\approx 0.55/M_\pi{}^2. \tag{35}$$

Actually, the narrow-resonance approximation is very misleading. Direct numerical evaluation of I, using the experimental (3,3) phase shift,[16] gives $I=0.63$, so that actually

$$V_1^{(+)}(0)|_{\text{magnetic dipole}}\approx 0.39/M_\pi{}^2. \tag{36}$$

In other words, the narrow-resonance approximation overestimates the integral I by 60%. [The narrow-resonance approximation is also misleading when used to evaluate the g_A sum rule. If only the (3,3) contribution to this sum rule is kept, one gets $g_A\approx 1.4$ when the integral is evaluated using the experimental πN cross section,[17] and $g_A=3$ when one uses the narrow-resonance approximation.] To sum up, the narrow-resonance approximation for the $N^*(1238)$ is useful for making order-of-magnitude estimates, but should be avoided in quantitative tests of sum rules.

B. Resonant Contributions

We turn next to the evaluation of the resonant contributions to $V_1^{(+)}(0)$ and $V_6^{(-)}(0)$, using Walker's photoproduction analysis. Walker[5] has parametrized each resonant multipole $\mathfrak{M}$ around the $N^*(1238)$ and

[15] G. F. Chew, F. E. Low, M. L. Goldberger, and Y. Nambu, Phys. Rev. **106**, 1337 (1957).

[16] We obtained the same numerical result using the (3,3) phase-shift parametrizations of Schmidt (Ref. 4) and of L. D. Roper, University of California Report No. UCRL-7846 (unpublished). For an independent evaluation of this integral, see D. Lyth, Phys. Letters **21**, 338 (1966).

[17] See W. I. Weisberger, Phys. Rev. Letters **14**, 1047 (1965); S. L. Adler, *ibid.* **14**, 1051 (1965).

TABLE II. Multipole contributions.

Multipole	Contribution to $[g_r/g_r(0)]V_1^{(+)}(0)$ (units of M_π^{-2})	Contribution to $[g_r/g_r(0)]V_6^{(-)}(0)$ (units of M_π^{-3})
$E_{0+}^{(1/2)}$	+0.055	+0.0329
$E_{0+}^{(3/2)}$	+0.081	−0.0212
$L_{0+}^{(1/2)}$		−0.0365
$L_{0+}^{(3/2)}$		+0.0238
$M_{1+}^{(3/2)}$	+0.413	+0.0133
$E_{1+}^{(3/2)}$	−0.088	+0.0471
$L_{1+}^{(3/2)}$		−0.0333
$M_{2-}^{(1/2)}$	−0.031	+0.0018
$E_{2-}^{(1/2)}$	+0.042	−0.0305
$L_{2-}^{(1/2)}$		+0.0281
Total	+0.472	+0.0255
PCAC prediction	+0.550	$\frac{g_r}{M_N}\left[\frac{g_A'(0)}{g_A(0)}-F_1{}^{V\prime}(0)\right]M_\pi{}^3 \approx 2M_\pi{}^2\left[\frac{g_A'(0)}{g_A(0)}+\frac{0.045}{M_\pi{}^2}\right]$

the $N^{**}(1520)$ in the form[18]

$$\mathfrak{M}=\frac{8\pi W}{M_N}\frac{A(|\mathbf{q}|_R/|\mathbf{q}|)^2\Gamma/2}{W_R-W-i\Gamma/2},$$
$$\Gamma=\Gamma_R\left(\frac{|\mathbf{q}|}{|\mathbf{q}|_R}\right)^3\frac{1+0.7735|\mathbf{q}|_R{}^2/M_\pi{}^2}{1+0.7735|\mathbf{q}|^2/M_\pi{}^2}. \tag{37}$$

The parameters A, Γ_R, W_R, and $|\mathbf{q}|_R$ are listed in Table I. Using Eqs. (37) and (31) we have calculated the integrals for $V_1^{(+)}(0)$ and $V_6^{(-)}(0)$, obtaining the results listed in Table II.

We note, first of all, that according to Table II the $M_{1+}^{(3/2)}$ contribution to $V_1^{(+)}(0)$ is

$$V_1^{(+)}(0)|_{\text{magnetic dipole}}\approx 0.41/M_\pi{}^2, \tag{38}$$

in good agreement with the value of $0.39/M_\pi{}^2$ obtained above from the CGLN–Schmidt-Höhler work. The multipole $E_{1+}^{(3/2)}$ makes a significant contribution to the sum rule because it appears in Eq. (28a) with a coefficient three times as large as the coefficient of $M_{1+}^{(3/2)}$. Walker's Im$E_{1+}^{(3/2)}$ has a constant negative sign across the $N^*(1238)$. If the suggestion of the CGLN model[14] [that Im$E_{1+}^{(3/2)}$ changes sign from negative to positive around the (3,3) resonance peak] should prove to be correct, then the value for the $E_{1+}^{(3/2)}$ contribution given in Table II may be an overestimate.

Looking at the contributions to $V_6^{(-)}(0)$, it may at first seem surprising that the small $E_{1+}^{(3/2)}$ multipole makes a much bigger contribution than the large $M_{1+}^{(3/2)}$ multipole. But a glance at Eq. (28b) shows the reason why—the ratio of the coefficients of $M_{1+}^{(3/2)}$

[18] Equation (37) does not give the multipoles M_{2-}, E_{2-} the correct threshold behavior, but the $N^{**}(1520)$ is far enough from threshold so that this is not too important. To make the off-mass-shell correction we have multiplied each $\mathfrak{M}$ by $[|\mathbf{q}|_{M_\pi{}^f=0}/|\mathbf{q}|_{M_\pi{}^f=M_\pi}]$, so that the off-mass-shell multipoles all have the correct threshold behavior.

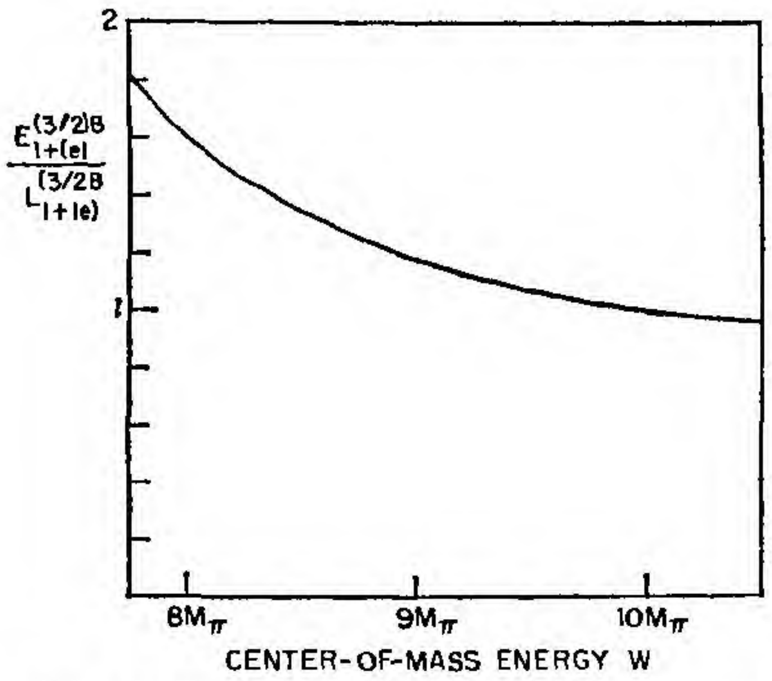


FIG. 2. Ratio of electric to longitudinal multipoles, for $k^2=0$.

and $E_{1+}^{(3/2)}$ in the integral for $V_6^{(-)}(0)$ is

$$\frac{1}{W+M_N}\left[\frac{-3(3W+M_N)}{W^2-M_N^2}\right]^{-1}=-\frac{(W-M_N)}{3(3W+M_N)}, \quad (39)$$

which is numerically ≈ -0.02 at the peak of the $N^*(1238)$. In other words, the $M_{1+}^{(3/2)}$ contribution is very strongly kinematically suppressed. The longitudinal multipoles contribute with strength comparable to the electric multipoles. To emphasize the dubious nature of the approximation of Eq. (31) for the longitudinal multipoles, we have computed the Born approximations $E_{1+}^{(3/2)}$ and $L_{1+}^{(3/2)}$ from the diagrams of Fig. 1, splitting them into parts proportional to the electric charge e and the difference of the nucleon total magnetic moments $\mu_p-\mu_n$:

$$\begin{aligned} E_{1+}^{(3/2)B}&=eE_{1+(e)}^{(3/2)B}+(\mu_p-\mu_n)E_{1+(\mu)}^{(3/2)B},\\ L_{1+}^{(3/2)B}&=eL_{1+(e)}^{(3/2)B}+(\mu_p-\mu_n)L_{1+(\mu)}^{(3/2)B}. \end{aligned} \quad (40)$$

(Numerically, the e terms, which come largely from the pion-exchange graph, are much larger than the μ terms, which come from the crossed nucleon graph.) One can verify, by direct calculation, that at $|\mathbf{k}|=0$ (for all $k^2\neq 0$),

$$\frac{E_{1+(e)}^{(3/2)B}}{L_{1+(e)}^{(3/2)B}}=\frac{E_{1+(\mu)}^{(3/2)B}}{L_{1+(\mu)}^{(3/2)B}}=1. \quad (41)$$

But at the physical threshold $|\mathbf{q}|=0$ we find for real photons that

$$\begin{aligned} E_{1+(e)}^{(3/2)B}/L_{1+(e)}^{(3/2)B}&=1.88,\\ E_{1+(\mu)}^{(3/2)B}/L_{1+(\mu)}^{(3/2)B}&=0. \end{aligned} \quad (42)$$

In Fig. 2 we have plotted the ratio of the numerically dominant e terms as a function of energy. Clearly, in determining the longitudinal multipole contributions to $V_6^{(-)}(0)$, assumptions such as Eq. (31) are unreliable and there will be no substitute for measurement of the longitudinal multipoles in electroproduction experiments.

C. Nonresonant S Wave

It is well known that there is an important s-wave contribution to charged-pion photoproduction. Since the s-wave pion-nucleon phase shifts are of order 15°–20° in the low-energy region, the imaginary parts of the s-wave amplitudes will make an important contribution to the integrals for $V_1^{(+)}(0)$ and $V_6^{(-)}(0)$. We estimate this contribution as follows. The Born approximations for the s-wave multipoles $E_{0+}^{(\pm,0)}$ are[14]

$$E_{0+}^{(+)B}\approx-\frac{W-M_N}{M_N}\frac{4.70}{M_\pi}, \quad E_{0+}^{(0)B}\approx-\frac{W-M_N}{M_N}\frac{0.88}{M_\pi}, \quad (43)$$
$$E_{0+}^{(-)B}\approx\frac{1}{M_\pi}\left[1+\frac{1-V^2}{2V}\ln\left(\frac{1+V}{1-V}\right)\right], \quad V=|\mathbf{q}|/q_0.$$

Pion-photoproduction experiments, as analyzed by Schmidt,[4] indicate that (i) in charged-pion photoproduction, the multipole E_{0+} is equal to the Born approximation; (ii) in neutral-pion photoproduction, $(M_N/W)E_{0+}$ is independent of energy, and at threshold is roughly one-half of the Born approximation. The charged and neutral pion amplitudes are related to $E_{0+}^{(\pm,0)}$ by

$$\begin{aligned} E_{0+}^{(\pi^+)}&=(1/\sqrt{2})(E_{0+}^{(-)}+E_{0+}^{(0)}),\\ E_{0+}^{(\pi^0)}&=\tfrac{1}{2}[E_{0+}^{(+)}+E_{0+}^{(0)}]. \end{aligned} \quad (44)$$

If we assume that the isoscalar amplitude (which is small anyway) is not much different from its Born approximation,[19] then the experimental results imply

$$\begin{aligned} \mathrm{Re}E_{0+}^{(+)}&\approx-\frac{W}{M_N+M_\pi}0.4\frac{4.70}{M_N},\\ \mathrm{Re}E_{0+}^{(-)}&\approx E_{0+}^{(-)B}. \end{aligned} \quad (45)$$

We get the imaginary parts of the multipoles $E_{0+}^{(1/2,3/2)}$ by using the Fermi-Watson theorem, which tells us that

$$\begin{aligned} \mathrm{Im}E_{0+}^{(1/2)}&\approx\sin\delta_1\,\mathrm{Re}E_{0+}^{(1/2)}\\ &=\sin\delta_1[\mathrm{Re}E_{0+}^{(+)}+2\,\mathrm{Re}E_{0+}^{(-)}],\\ \mathrm{Im}E_{0+}^{(3/2)}&\approx\sin\delta_3\,\mathrm{Re}E_{0+}^{(3/2)}\\ &=\sin\delta_3[\mathrm{Re}E_{0+}^{(+)}-\mathrm{Re}E_{0+}^{(-)}], \end{aligned} \quad (46)$$

with δ_1, δ_3 the $I=\frac{1}{2}$, $\frac{3}{2}$ s-wave pion-nucleon phase shifts.

The numbers given in Table II have been obtained by using Eqs. (45) and (46), integrating from threshold to a center-of-mass energy $W=10\ M_\pi$, and taking the the pion-nucleon phase shifts from Roper's $l_m=4$ analysis.[20] Adding the s-wave result to the other

[19] This is suggested by the CGLN model, in which the isoscalar amplitude is given by the Born approximation, but the isovector amplitude differs appreciably from the Born approximation due to the presence of the dispersion integral over the $N^*(1238)$.

[20] L. D. Roper, Ref. 16.

contributions to $V_1^{(+)}(0)$ raises the total to about 0.85 of the value predicted by PCAC.[21] If we assume Eq. (31) for L_{0+}, the result for $V_6^{(-)}(0)$ obtained from the resonant multipoles is changed very little.

[21] If $\mathrm{Re}E_{0+}^{(+)}$ is taken to be zero, the $E_{0+}^{(1/2)}$ and $E_{0+}^{(3/2)}$ contributions to $V_1^{(+)}(0)$ become $0.062/M_\pi^2$ and $0.066/M_\pi^2$, respectively. Thus, as expected, the s-wave contribution comes mainly from the charged-pion photoproduction amplitude $E_{0+}^{(-)}$. The only multipole significant in the low-energy region which we have omitted from our analysis is M_{1-}. While $\mathrm{Re}M_{1-}$ is known, the P_{11} pion-nucleon phase shift becomes large only when the inelasticity in this channel is also large. This means that $\mathrm{Im}M_{1-}$ cannot then be reliably determined by the Fermi-Watson theorem.

ACKNOWLEDGMENTS

We wish to thank Professor R. L. Walker for supplying us with his fits to the photoproduction data. One of us (S. L. A.) wishes to acknowledge a very pleasant visit at the California Institute of Technology, where part of this work was done.

PHYSICAL REVIEW VOLUME 169, NUMBER 5 25 MAY 1968

Possible Measurement of the Nucleon Axial-Vector Form Factor in Two-Pion Electroproduction Experiments

STEPHEN L. ADLER
Institute for Advanced Study, Princeton, New Jersey 08540

AND

WILLIAM I. WEISBERGER*†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey 08540

(Received 8 December 1967)

Current-algebra techniques and the hypothesis of partially conserved axial-vector current are used to derive a low-energy theorem for the reaction $e+N\rightarrow e+N+\pi+\pi$(soft). Particular attention is paid to satisfying the requirements of gauge covariance. Except for recoil corrections, the resulting matrix element is proportional to the nucleon axial-vector form factors, and we suggest that this electromagnetic process may be used to measure $g_A(k^2)$.

I. INTRODUCTION

MEASUREMENT of the momentum-transfer dependence of the nucleon axial-vector form factor $g_A(k^2)$ would clearly be of great interest, since it would give information about the spectrum of axial-vector mesons, just as our experimental knowledge of the nucleon electromagnetic form factors has provided much useful information about the vector mesons. Unfortunately, the elastic and inelastic weak-interaction experiments[1] to measure $g_A(k^2)$ are much more difficult than their electromagnetic counterparts, and as a result very little about $g_A(k^2)$ is known at present. Clearly, it would be useful to have alternative, even if very indirect, methods of measuring $g_A(k^2)$. We discuss in this paper the possibility of measuring $g_A(k^2)$ in the electroproduction reaction

$$e+N\rightarrow e+N+\pi+\pi(\text{soft}), \quad (1)$$

assuming the validity of the current algebra and of the partially conserved axial-vector current (PCAC) hypotheses. This possibility is suggested by the recent work of a number of authors,[2] showing that when current-algebra–PCAC methods are applied to the photoproduction reaction $\gamma+N\rightarrow N+\pi+\pi$(soft), which is the $k^2=0$ case of Eq. (1), the results of the old Cutkosky-Zachariasen static model[3] are obtained, with the dominant term coming from the matrix element of the axial-vector current $\langle N\pi|J_\lambda{}^A|N\rangle$.

* A. P. Sloan Foundation Fellow.

† Supported in part by the U. S. Air Force Office of Research, Air Research and Development Command, under Contract No. AF 49 (638)-1545.

[1] For a discussion of the determination of $g_A(k^2)$ in neutrino experiments, see E. C. M. Young, CERN Report 67-12 (unpublished).

[2] T. Ebata, Phys. Rev. 154, 1341 (1967); P. Carruthers and H. W. Huang, Phys. Letters 24B, 464 (1967); P. Narayanaswamy and B. Renner, Nuovo Cimento 53A, 107 (1968); S. M. Berman (unpublished) (Berman has also considered the extension to electroproduction); W. I. Weisberger (unpublished).

[3] R. E. Cutkosky and F. Zachariasen, Phys. Rev. 103, 1108 (1956). [See also P. Carruthers and H. Wong, *ibid.* 128, 2382 (1962).] The soft-pion result generalizes their model to a relativistic framework in the same way that Chew, Goldberger, Low, and Nambu extended the Chew-Low static model for $N_{3,3}{}^*$ photoproduction.

In Sec. II we apply soft-pion methods to the reaction of Eq. (1), and get a relation between the matrix element for this process and the matrix elements for single-pion weak production and electroproduction, $\langle N\pi|J_\lambda{}^A|N\rangle$ and $\langle N\pi|J_\lambda{}^{\text{EM}}|N\rangle$. By carefully keeping all pion pole diagrams, we eliminate some discrepancies noted in the previous work on two-pion photoproduction. The matrix element $\langle N\pi|J_\lambda{}^A|N\rangle$ can be related, in turn, to the axial-vector form factor $g_A(k^2)$, using models analogous to the very successful CGLN[4] treatment of pion photoproduction.

In Sec. III we retain only the $I=J=\frac{3}{2}$ partial wave, treated in the CGLN approximation, and discuss the possibility of measuring $g_A(k^2)$ in the reaction $e+N\rightarrow N_{3,3}{}^*(1238)+\pi$(soft).

II. DERIVATION

We will consider the electroproduction reaction

$$e(k_1)+N(p_1)\rightarrow e(k_2)+N(p_2)+\pi(q)+\pi^s(q_s), \quad (2)$$

with the superscript s an isospin index. Letting $k=k_1-k_2$ be the four-momentum transfer between the electrons, the hadronic matrix element for Eq. (2) is

$$M_\lambda=\int d^4x d^4y\, e^{ik\cdot x}e^{-iq_s\cdot y}(-\Box_y{}^2+M_\pi{}^2)\times\langle N(p_2)\pi(q)|T(\phi_{\pi^s}(y)J_\lambda{}^{\text{EM}}(x))|N(p_1)\rangle. \quad (3)$$

We wish to find the limit of Eq. (3) when π^s is soft, that is, as $q_s\rightarrow 0$. This can be done by the standard soft-pion methods[5]; the only delicate point is to insure that our soft-pion approximation for M_λ satisfies gauge invariance.

Let us begin then by studying the gauge properties of M_λ. Multiplying Eq. (3) by $-ik_\lambda$, integrating by parts

[4] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. 106, 1345 (1957). Hereafter referred to as CGLN.

[5] See, for example, S. L. Adler and F. J. Gilman, Phys. Rev. 152, 1460 (1966).

with respect to x, and using $\partial_\lambda J_\lambda^{\rm EM}=0$ gives

$$-ik_\lambda M_\lambda=\int d^4xd^4y\, e^{ik\cdot x}e^{-iq_s\cdot y}(-\Box_y^2+M_\pi^2)$$
$$\times\langle N(p_2)\pi(q)\,|\,\delta(x_0-y_0)[J_0^{\rm EM}(x),\phi_{\pi^s}(y)]\,|N(p_1)\rangle. \quad (4)$$

In all simple canonical field theories involving pions one finds[6]

$$\delta(x_0-y_0)[J_0^{\rm EM}(x),\phi_{\pi^s}(y)]=i\epsilon_{3sc}\delta^4(x-y)\phi_{\pi^c}(y); \quad (5)$$

substituting this into Eq. (4) and finally integrating by parts with respect to y gives

$$k_\lambda M_\lambda=-\epsilon_{3sc}(q_s^2+M_\pi^2)\int d^4y$$
$$\times e^{i(k-q_s)\cdot y}\langle N(p_2)\pi(q)\,|\,\phi_{\pi^c}(y)\,|N(p_1)\rangle$$
$$=-\frac{\epsilon_{3sc}(q_s^2+M_\pi^2)}{(k-q_s)^2+M_\pi^2}\int d^4y$$
$$\times e^{i(k-q_s)\cdot y}\langle N(p_2)\pi(q)\,|\,J_{\pi^c}(y)\,|N(p_1)\rangle. \quad (6)$$

As expected, when π^s is on the mass shell, $k_\lambda M_\lambda=0$, but in the off-shell case the divergence of M_λ is nonzero. Our soft-pion approximation for M_λ will not actually satisfy Eq. (6) exactly, but will obey the approximate version

$$k_\lambda M_\lambda\approx-\frac{\epsilon_{3sc}(q_s^2+M_\pi^2)}{(k-q_s)^2+M_\pi^2}\int d^4y$$
$$\times e^{ik\cdot y}\langle N(p_2)\pi(q)\,|\,J_{\pi^c}(y)\,|N(p_1)\rangle, \quad (7)$$

obtained by neglecting q_s in the matrix element of J_π but keeping q_s in the rapidly varying factor $(q_s^2+M_\pi^2)/[(k-q_s)^2+M_\pi^2]$. Clearly, Eqs. (6) and (7) are identical both in the soft-pion limit ($q_s=0$) and on the mass shell ($q_s^2=-M_\pi^2$).

In applying PCAC to Eq. (3), it is helpful to introduce the "proper part" J_λ^{AP} of the axial-vector current, defined as follows: Let a and b be arbitrary hadron states, and let $q=p_a-p_b$. Then we define J_λ^{AP} by

$$\langle a|J_\lambda^{AP}|b\rangle=\langle a|J_\lambda^A|b\rangle+\frac{q_\lambda}{M_\pi^2}\langle a|q_\sigma J_\sigma^A|b\rangle, \quad (8)$$

which implies that

$$\langle a|J_\lambda^A|b\rangle=\langle a|J_\lambda^{AP}|b\rangle-\frac{q_\lambda}{q^2+M_\pi^2}\langle a|q_\sigma J_\sigma^{AP}|b\rangle, \quad (9)$$

$$\langle a|q_\lambda J_\lambda^{AP}|b\rangle=\frac{q^2+M_\pi^2}{M_\pi^2}\langle a|q_\lambda J_\lambda^A|b\rangle. \quad (10)$$

Clearly, the proper current J_λ^{AP} has no pion pole; Eq. (9) is thus a convenient decomposition of the axial-vector current into pion-pole and non-pion-pole pieces. [As an illustration, let us take a and b to be nucleons. Then $\langle N|J_\lambda^A|N\rangle\propto\bar u(g_A\gamma_\lambda\gamma_5+iq_\lambda h_A\gamma_5)u$. In the approximation in which the induced pseudoscalar form factor h_A is given by $h_A=2M_Ng_A/(q^2+M_\pi^2)$, the proper part of $\langle N|J_\lambda^A|N\rangle$ is just the piece $\bar u g_A\gamma_\lambda\gamma_5 u$.] Let us now introduce the PCAC hypothesis in the form

$$\partial_\sigma J_\sigma^{sA}=\frac{M_NM_\pi^2g_A}{g_r(0)}\phi_{\pi^s}. \quad (11)$$

Then using Eq. (10) we can write Eq. (11) as

$$\partial_\sigma J_\sigma^{sAP}=\frac{M_Ng_A}{g_r(0)}J_{\pi^s}, \quad (12)$$

which says that the divergence of the proper part of the axial-vector current is a smooth interpolating operator for the pion source. Thus, we can rewrite the gauge condition [Eq. (7)] in the alternative form

$$k_\lambda M_\lambda\approx\frac{i\epsilon_{3sc}(q_s^2+M_\pi^2)k_\sigma}{(k-q_s)^2+M_\pi^2}\frac{g_r(0)}{M_Ng_A}\int d^4y$$
$$\times e^{ik\cdot y}\langle N(p_2)\pi(q)\,|\,J_\sigma^{cAP}(y)\,|N(p_1)\rangle. \quad (13)$$

To get a soft-pion approximation for M_λ, we substitute Eq. (11) into Eq. (3) and integrate by parts with respect to y. This gives

$$\frac{M_\pi^2}{q_s^2+M_\pi^2}M_\lambda=M_\lambda^{\rm ETC}+M_\lambda^{\rm SURF}, \quad (14)$$

with

$$M_\lambda^{\rm ETC}=-i\epsilon_{s3c}\frac{g_r(0)}{M_Ng_A}\int d^4x$$
$$\times e^{i(k-q_s)\cdot x}\langle N(p_2)\pi(q)\,|\,J_\lambda^{cA}(x)\,|N(p_1)\rangle \quad (15)$$

the equal-time commutator of J_0^{sA} with $J_\lambda^{\rm EM}$,[7] and with

$$M_\lambda^{\rm SURF}=iq_{s\sigma}\frac{g_r(0)}{M_Ng_A}\int d^4xd^4y\, e^{ik\cdot x}e^{-iq_s\cdot y}$$
$$\times\langle N(p_2)\pi(q)\,|\,T(J_\sigma^{sA}(y)J_\lambda^{\rm EM}(x))\,|N(p_1)\rangle \quad (16)$$

the remainder. Separating Eq. (15) for $M_\lambda^{\rm ETC}$ into a

[6] When integrated over space with respect to x, Eq. (5) becomes $[I_3+\frac{1}{2}Y,\phi_{\pi^s}]=i\epsilon_{3sc}\phi_{\pi^c}$, which is just the statement that the pion is a particle with the quantum numbers $I=1$, $Y=0$. The local form, Eq. (5), follows from the integrated version in canonical field theories, since in such theories the charge density $J_0^{\rm EM}$ is a bilinear form in the canonical fields and momenta, and thus $[J_0^{\rm EM}(x),\phi_{\pi^s}(y)]|_{x_0=y_0}$ contains no gradient of δ-function terms which vanish when integrated spatially.

[7] We have, of course, evaluated the equal-time commutator using the Gell-Mann algebra of currents [M. Gell-Mann, Physics 1, 63 (1964)]. The possible presence of Schwinger terms in the time-space commutators is irrelevant because of the cancellation of the Schwinger term and "seagull-diagram" contributions in soft-pion calculations. See, for example, S. L. Adler and R. F. Dashen, *Current Algebras* (W. A. Benjamin, Inc., New York, 1968), Chap. 3.

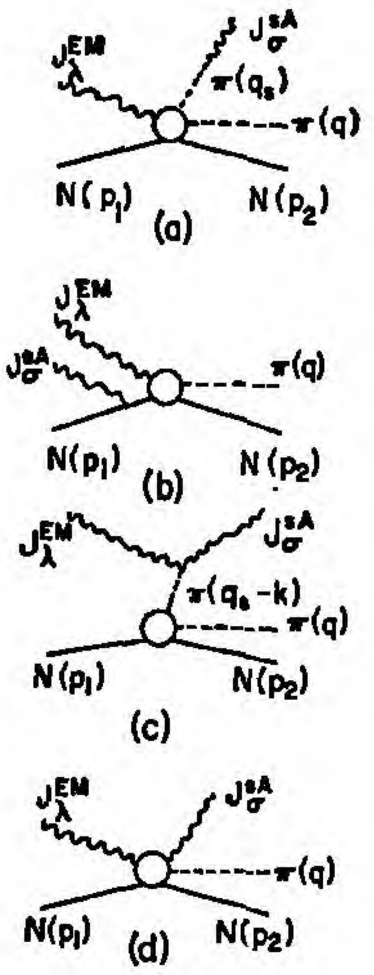


FIG. 1. Contributions to $M_\lambda^{\rm SURF}$ [Eq. (16)]. (a) Axial-vector current couples to a virtual pion. (b) Axial-vector current attaches to the initial external nucleon line in single-pion electroproduction. There is a similar diagram (not shown) in which the axial-vector current attaches to the final external nucleon line. (c) Axial-vector current and vector current attach to a virtual pion at the same space-time point [a "seagull" diagram]. (d) Axial-vector current couples to internal lines in the matrix element $\langle N(p_2)\pi(q)| \times T(J_\sigma^{sA}(y)J_\lambda^{\rm EM}(x))|N(p_1)\rangle$.

proper part and a remainder gives

$$M_\lambda^{\rm ETC}=-i\epsilon_{s3c}\frac{g_r(0)}{M_Ng_A}\bigg[\int d^4x\, e^{i(k-q_s)\cdot x}\langle N(p_2)\pi(q)| \times J_\lambda^{cAP}(x)|N(p_1)\rangle-\frac{(k-q_s)_\lambda(k-q_s)_\eta}{(k-q_s)^2+M_\pi^2}\int d^4x \times e^{i(k-q_s)\cdot x}\langle N(p_2)\pi(q)|J_\eta^{cAP}(x)|N(p_1)\rangle\bigg] \tag{17}$$

$$\approx -i\epsilon_{s3c}\frac{g_r(0)}{M_Ng_A}\bigg[\int d^4x\, e^{ik\cdot x}\langle N(p_2)\pi(q)|J_\lambda^{cAP}(x) \times|N(p_1)\rangle-\frac{(k-q_s)_\lambda k_\eta}{(k-q_s)^2+M_\pi^2}\int d^4x\, e^{ik\cdot x} \times\langle N(p_2)\pi(q)|J_\eta^{cAP}(x)|N(p_1)\rangle\bigg]. \tag{18}$$

In going from Eq. (17) to Eq. (18) we have neglected q_s in matrix elements of the proper part J_λ^{cAP} and its divergence $\partial_\eta J_\eta^{cAP}$, but have retained q_s in the rapidly varying factor $(k-q_s)_\lambda/[(k-q_s)^2+M_\pi^2]$. The surface term $M_\lambda^{\rm SURF}$ contains four types of terms, shown in Figs. 1(a)–1(d). In Fig. 1(a), the axial-vector current couples to a virtual pion; it is easy to see that

$$M_\lambda^{\rm SURF(a)}=\frac{-q_s^2}{q_s^2+M_\pi^2}M_\lambda. \tag{19}$$

In Fig. 1(b), the axial current attaches to an external nucleon line in single-pion electroproduction; an expression for $M_\lambda^{\rm SURF(b)}$ can be obtained from the usual axial-current insertion rules and is given below. In Fig. 1(c), the axial current and vector current attach to a virtual pion at the same space-time point; this is a "seagull" diagram contributing to virtual radiative pion decay and may be calculated to be

$$M_\lambda^{\rm SURF(c)}=\epsilon_{s3c}q_{s\sigma}\frac{1}{(k-q_s)^2+M_\pi^2}\int d^4x \times e^{i(k-q_s)\cdot x}\langle N(p_2)\pi(q)|J_{\pi^c}(x)|N(p_1)\rangle\delta_{\lambda\sigma} \tag{20}$$

$$\approx -i\epsilon_{s3c}\frac{g_r(0)}{M_Ng_A}\frac{q_{s\lambda}k_\eta}{(k-q_s)^2+M_\pi^2}\int d^4x \times e^{ik\cdot x}\langle N(p_2)\pi(q)|J_\eta^{cAP}(x)|N(p_1)\rangle. \tag{21}$$

Finally, in Fig. 1(d) the axial current couples to internal lines in the matrix element

$$\langle N(p_2)\pi(q)|T(J_\sigma^{sA}(y)J_\lambda^{\rm EM}(x))|N(p_1)\rangle;$$

consequently, $M_\lambda^{\rm SURF(d)}$ is of order q_s and may be neglected.

Comparing Eq. (21) with Eq. (18), we see that the effect of including the radiative pion decay diagram is to change the coefficient of the pion-pole term in $M_\lambda^{\rm ETC}$ from $(k-q_s)_\lambda$ to $(k-2q_s)_\lambda$. This eliminates the factor of two discrepancy noted by Carruthers and Huang,[2] who neglected $M_\lambda^{\rm SURF(c)}$, and leads to the satisfaction of the approximate gauge condition (13).

Combining all the terms, we may write our answer as follows:

$$M_\lambda=(2\pi)^4\delta^4(p_2+q-p_1-k)\left(\frac{M_N^2}{2p_{10}p_{20}q_0}\right)^{1/2} \times\bar u(p_2)N_\lambda u(p_1)+O(q_s), \tag{22}$$

$$N_\lambda=\epsilon_{s3c}\left[\frac{(2q_s-k)_\lambda k_\eta}{(k-q_s)^2+M_\pi^2}+\delta_{\lambda\eta}\right]\left(\frac{-ig_r(0)}{M_Ng_A}\right)O_\eta^{cAP} +\frac{g_r(0)}{2M_N}\tau^s q_s\cdot\gamma_5\frac{p_2+iM_N}{2p_2\cdot q_s}O_\lambda^{\rm EM} +O_\lambda^{\rm EM}\frac{p_1+iM_N}{-2p_1\cdot q_s}\frac{g_r(0)}{2M_N}\tau^s q_s\cdot\gamma_5,$$

where O_η^{cAP} and $O_\lambda^{\rm EM}$ are defined by

$$\langle N(p_2)\pi(q)|J_\eta^{cAP}|N(p_1)\rangle=\left(\frac{M_N^2}{2p_{10}p_{20}q_0}\right)^{1/2} \times\bar u(p_2)O_\eta^{cAP}u(p_1), \tag{23a}$$

$$\langle N(p_2)\pi(q)|J_\lambda^{\rm EM}|N(p_1)\rangle=\left(\frac{M_N^2}{2p_{10}p_{20}q_0}\right)^{1/2} \times\bar u(p_2)O_\lambda^{\rm EM}u(p_1). \tag{23b}$$

The terms proportional to $O_\lambda^{\rm EM}$ are the single-pion electroproduction contribution $M_\lambda^{\rm SURF(b)}$ mentioned above.[8] Since the single-pion electroproduction matrix

[8] In writing the matrix element $O_\lambda^{\rm EM}$ we neglect the additional momentum q_s carried by the intermediate nucleon. It is clear that the error is $O(q_s)$, consistent with our approximation.

element is gauge-invariant, we have $k_\lambda(p_2+iM_N)\times O_\lambda^{\mathrm{EM}}u(p_1)=k_\lambda\bar{u}(p_2)O_\lambda^{\mathrm{EM}}(p_1+iM_N)=0$, and thus the divergence of N_λ is

$$k_\lambda N_\lambda=\epsilon_{s3c}\frac{q_s^2+M_\pi^2}{(k-q_s)^2+M_\pi^2}k_\eta\left(\frac{-ig_r(0)}{M_N g_A}\right)O_\eta^{cAP}. \quad (24)$$

Combining Eqs. (22)–(24), it is clear that the approximate gauge condition of Eq. (13) is satisfied. In particular, when $q_s^2=-M_\pi^2$, $k_\lambda N_\lambda=0$, so on-mass-shell Eq. (22) gives a gauge-invariant approximation to the matrix element for two-pion electroproduction.

III. DISCUSSION

Let us now briefly consider the possibility of indirectly measuring $g_A(k^2)$ in the reaction $e+N\to e+N+\pi+\pi(\mathrm{soft})$, by use of Eqs. (22)–(23). For simplicity, we will restrict ourselves to the case in which the soft pion is at rest (threshold) in the center-of-mass frame of the final baryons,[9] and in which the hard pion and nucleon emerge in the (3,3) resonance. At the soft-pion threshold, the kinematic structure of two-pion electroproduction becomes identical to the kinematic structure of the more familiar case of single-pion electroproduction; this makes it easy to compute the two-pion cross section from the matrix element in Eqs. (22)–(23). When the hard π and N form an $N_{3,3}^*$, the matrix elements in Eqs. (23a) and (23b) describe weak production of the (3,3) resonance from a nucleon target and have been extensively studied.[10] The vector matrix element [Eq. (23b)] is found to be dominated by the magnetic dipole[11] amplitude $M_{1+}^{(3/2)}$, while the axial-vector matrix element [Eq. (23a)] is dominated by the electric, longitudinal, and scalar amplitudes $\mathcal{E}_{1+}^{(3/2)}$, $\mathcal{L}_{1+}^{(3/2)}$, and $\mathcal{H}_{1+(g_A)}^{(3/2)}$. [The subscript (g_A) indicates that the part of $\mathcal{H}_{1+}^{(3/2)}$ proportional to the induced pseudoscalar form factor h_A is to be dropped and only the part proportional to the axial-vector form factor g_A retained; this restriction arises because only the *proper* part of the axial-vector current appears in Eq. (23).] For momentum transfers k^2 less than 50 F^{-2}, a model which should give a good approximation to $M_{1+}^{(3/2)}$, $\cdots$ is

$$\begin{aligned}
M_{1+}^{(3/2)}&=M_{1+}^{(3/2)B}f_{1+}^{(3/2)}/f_{1+}^{(3/2)B},\\
\mathcal{E}_{1+}^{(3/2)}&=\mathcal{E}_{1+}^{(3/2)B}f_{1+}^{(3/2)}/f_{1+}^{(3/2)B},\\
\mathcal{L}_{1+}^{(3/2)}&=\mathcal{L}_{1+}^{(3/2)B}f_{1+}^{(3/2)}/f_{1+}^{(3/2)B},\\
\mathcal{H}_{1+(g_A)}^{(3/2)}&=\mathcal{H}_{1+(g_A)}^{(3/2)B}f_{1+}^{(3/2)}/f_{1+}^{(3/2)B},
\end{aligned} \quad (25)$$

where $f_{1+}^{(3/2)}$ is the pion-nucleon scattering amplitude in the (3,3) channel and where the superscript B denotes "Born approximation." Expressions for $f_{1+}^{(3/2)B}$, $M_{1+}^{(3/2)B}$, $\mathcal{E}_{1+}^{(3/2)B}$, $\cdots$ are given in the Appendix.[10]

[9] That is, the frame defined by $\mathbf{p}_2+\mathbf{q}+\mathbf{q}_s=0$. In the case $\mathbf{q}_s=0$ which we consider, the center-of-mass frame of the final baryons is identical with the center-of-mass frame of the hard pion and nucleon (the $N_{3,3}^*$ rest frame).

[10] S. L. Adler (to be published).

[11] Our multipoles are a factor $(8\pi W/M_N e)$ times those of Ref. 4.

TABLE I. Isospin coefficients.

	a_1	a_2	a_3
$e+p\to e+\pi^+(\mathrm{soft})+N_{3,3}^{*0}$			
$N_{3,3}^{*0}\to p+\pi^-$	$-\frac{1}{6}$	$-\frac{1}{12}$	0
$\searrow n+\pi^0$	$-\frac{1}{3}(\sqrt{2})^{-1}$	$-\frac{1}{12}\sqrt{2}$	$\frac{1}{6}\sqrt{2}$
$e+p\to e+\pi^-(\mathrm{soft})+N_{3,3}^{*++}$			
$N_{33}^{*++}\to p+\pi^+$	$\frac{1}{2}$	0	$-\frac{1}{6}$

A straightforward calculation shows that, in terms of the weak (3,3) production multipoles, the cross section for $e+N\to e+N_{3,3}^*+\pi(\mathrm{threshold})$ is given by

$$\begin{aligned}
\sigma_1(k^2,W)&\equiv\frac{1}{|\mathbf{q}_s|}\frac{d^3\sigma[e+N\to e+N_{3,3}^*+\pi(\mathrm{soft})]}{dq_{s0}dk^2dW}\bigg|_{q_{s0}=M_\pi}\\
&=\frac{\alpha^2}{\pi^3}\frac{g_r(0)^2}{M_N^2}\frac{(W+M_\pi)^2}{W^2+(W+M_\pi)^2-M_\pi^2}\frac{|\mathbf{q}|}{(k_{10}^L)^2}\\
&\quad\times\left[\frac{1}{2k^2}\left(1+\frac{2k_{10}k_{20}-\frac{1}{2}k^2}{|\mathbf{k}|^2}\right)[|A|^2+3|B|^2+|C|^2]\right.\\
&\qquad\left.+\frac{4k_{10}k_{20}-k^2}{|\mathbf{k}|^2}|D|^2\right],
\end{aligned} \quad (26)$$

$$\begin{aligned}
A&=\frac{1}{g_A}a_1\mathcal{E}_{1+}^{(3/2)}+a_2\frac{|\mathbf{k}|}{p_{10}}M_{1+}^{(3/2)},\\
B&=a_2\frac{|\mathbf{k}|}{p_{10}}M_{1+}^{(3/2)},\\
C&=a_3\frac{|\mathbf{q}|}{p_{20}}M_{1+}^{(3/2)},\\
D&=\frac{1}{g_A}a_1\\
&\quad\times\frac{(k_0-2M_\pi)\mathcal{L}_{1+}^{(3/2)}+2M_\pi(|\mathbf{k}|/k_0)\mathcal{H}_{1+(g_A)}^{(3/2)}}{k^2+2M_\pi k_0},
\end{aligned} \quad (27)$$

where k_{10}^L is the laboratory-frame initial electron energy, where q_{s0} and all other noninvariant quantities refer to the center-of-mass frame of the final baryons, and where W is the invariant mass of the resonating pion and nucleon. Values of the isospin coefficients $a_{1,2,3}$ are given in Table I. For comparison, the cross section for the ordinary (3,3) electroproduction reaction $e+p\to e+N_{3,3}^{*+}$ is

$$\begin{aligned}
\sigma_2(k^2,W)&\equiv\frac{d^2\sigma(e+p\to e+N_{3,3}^{*+})}{dk^2dW}=\frac{\alpha^2}{3\pi}\frac{|\mathbf{q}|}{(k_{10}^L)^2}\frac{1}{2k^2}\\
&\quad\times\left(1+\frac{2k_{10}k_{20}-\frac{1}{2}k^2}{|\mathbf{k}|^2}\right)|M_{1+}^{(3/2)}|^2.
\end{aligned} \quad (28)$$

Looking at Table I, we see that the most promising reaction for the measurement of $g_A(k^2)$ is

$$e+p\to e+\pi^-(\text{soft})+N_{3,3}^{*++} \searrow p+\pi^+ \tag{29}$$

for the following three reasons: (1) The coefficient a_1 of the axial-vector multipoles is the largest in this case. (2) The coefficient a_2 vanishes and, consequently, the vector multipole $M_{1+}^{(3/2)}$ enters only through the very small recoil-correction term $|C|^2$. (3) In this case there is no soft-pion background coming from single-pion electroproduction, which can only lead to a soft π^+ or π^0.

Because the Born approximations $\mathcal{E}_{1+}^{(3/2)B}$ and $\mathcal{L}_{1+}^{(3/2)B}$ are known functions of W and k^2, and are proportional to $g_A(k^2)$, Eq. (26) [apart from the small term $|C|^2$] is proportional to $g_A(k^2)^2$, and thus a measurement of σ_1 as a function of k^2 will determine the momentum transfer dependence of g_A.[12]

There is, however, a possible problem, which may be illustrated by comparing Eq. (26) with Eq. (28) for ordinary (3,3) resonance electroproduction. Just as σ_1 is proportional to $g_A(k^2)^2$, σ_2 is proportional to $F^V(k^2)^2$, where $F^V(k^2)$ is an isovector electromagnetic form factor. There seems to be some evidence that the axial-vector form factor $g_A(k^2)$ falls off considerably more slowly with k^2 than does $F^V(k^2)$. This in turn suggests that the soft pion$+N_{3,3}^*$ production cross section σ_1 falls off much more slowly with k^2 than does the $N_{3,3}^*$ cross section σ_2. Unfortunately, however, this conclusion is not correct. The reason is that the multipoles $M_{1+}^{(3/2)}$ and $\mathcal{E}_{1+}^{(3/2)}$ have different small-$|\mathbf{k}|$ threshold behavior,

$$\left.\begin{array}{l} M_{1+}^{(3/2)}\sim|\mathbf{k}| \\ \mathcal{E}_{1+}^{(3/2)}\sim 1 \end{array}\right\} |\mathbf{k}|\to 0, \tag{30}$$

and this behavior, in the model of Eq. (25), persists into the physical region as well. As a result, the correct statement about the relative rates of decrease of σ_1 and σ_2 is that

$$\frac{\sigma_1(k^2)/\sigma_1(0)}{\sigma_2(k^2)/\sigma_2(0)}\approx\frac{[g_A(k^2)/g_A]^2}{[F^V(k^2)]^2}\frac{|\mathbf{k}|^2{}_{k^2=0}}{|\mathbf{k}|^2{}_{k^2}} \approx\frac{[g_A(k^2)/g_A]^2}{[F^V(k^2)]^2}\frac{(W-M_N)^2}{(W-M_N)^2+k^2}. \tag{31}$$

Even if $g_A(k^2)$ falls off appreciably more slowly than $F^V(k^2)$, the effect of the factor $(W-M_N)^2/[(W-M_N)^2+k^2]$ is to cause σ_1 to decrease more rapidly than σ_2.

The importance of the threshold behavior in Eq. (31) illustrates a problem which might invalidate Eq. (22), our soft-pion approximation for the two-pion production matrix element, and thus destroy the possibility of measuring $g_A(k^2)$ in the reaction Eq. (29). In deriving Eq. (22), we have neglected terms of first order or higher in the soft-pion four-momentum q_s. At $k^2=0$, we feel fairly justified in this approximation, since it leads to the Cutkosky-Zachariasen formulas, which seem to work. However, it is always possible that some of the terms of order q_s, which are negligible at $k^2=0$, increase rapidly relative to the terms of zeroth order in q_s as k^2 increases, because of a different threshold behavior in $|\mathbf{k}|$. If this happened, the soft-pion approximation could become bad precisely in the large-k^2 region, where we must look to measure $g_A(k^2)$. Hopefully, this does not happen, but in using Eq. (22) to interpret two-pion electroproduction experiments, this danger must be kept in mind. A more detailed investigation of this problem is being undertaken.

APPENDIX

We give here expressions for the Born approximations $f_{1+}^{(3/2)B}$, $M_{1+}^{(3/2)B}$, $\mathcal{E}_{1+}^{(3/2)B}$, $\mathcal{L}_{1+}^{(3/2)B}$, and $\mathcal{H}_{1+(g_A)}^{(3/2)B}$:

$$f_{1+}^{(3/2)B}=-\frac{g_r^2}{8\pi W|\mathbf{q}|^2}[W_-(p_{20}+M_N)A(\bar{a})+W_+(p_{20}-M_N)C(\bar{a})],$$

$$M_{1+}^{(3/2)B}=\frac{W^2|\mathbf{q}||\mathbf{k}|}{O_{2+}}\left(\frac{-g_r}{4M_N^2}\right)[F_1^V(k^2)+2M_NF_2^V(k^2)]\left[\frac{M_NW_-(p_{10}+M_N)}{W^2}\frac{A(a)}{|\mathbf{q}|^2|\mathbf{k}|^2}-\frac{W_+}{W^2}\frac{B(a)}{|\mathbf{q}||\mathbf{k}|}+\frac{M_NW_+}{W^2(p_{20}+M_N)}\frac{C(a)}{|\mathbf{q}||\mathbf{k}|}\right]+\text{nucleon and pion charge terms},$$

[12] A similar calculation would lead to a determination of $g_A(k^2)$ in electroproduction of a single soft pion. The relevant matrix elements are given in Ref. 5, which gives further references. Experimental data on single- and double-pion photoproduction reactions indicate that double-pion electroproduction may yield more reliable results for $g_A(k^2)$ than single-pion electroproduction. The reason is that the soft-pion matrix element seems to give an accurate description of the experimental results for two-pion photoproduction up to about 100 MeV above the $N_{3,3}^*+\pi$ threshold, while the single-pion photoproduction is dominated by $N_{3,3}^*$ production (which cannot be described by soft-pion methods) as soon as one goes away from threshold. In fact, it is interesting to note that the recent DESY results on $\gamma+p\to N_{3,3}^{*++}+\pi^-$ show a cross section rising less rapidly above threshold than indicated by earlier experiments and agree within experimental error with the prediction of the Cutkosky-Zachariasen model. The relevant experimental results and references are given in Fig. 9 of M. G. Hauser, Phys. Rev. 160, 1215 (1967). If both methods of measuring $g_A(k^2)$ are feasible, one will be happy to have two independent determinations.

$$\mathcal{E}_{1+}{}^{(3/2)B}=W^2O_{1+}|\mathbf{q}|\left(\frac{-g_rg_A(k^2)}{2M_N}\right)\left[\frac{\frac{1}{2}W_-(p_{10}-M_N)}{W^2}\frac{A(a)}{|\mathbf{q}|^2|\mathbf{k}|^2}+\frac{2}{W^2}\frac{B(a)}{|\mathbf{q}||\mathbf{k}|}\right.$$
$$\left.+\frac{\frac{1}{2}W_+}{W^2(p_{20}+M_N)}\frac{C(a)}{|\mathbf{q}||\mathbf{k}|}-\frac{3}{W^2}\frac{E(a)}{(p_{10}+M_N)(p_{20}+M_N)}\right],$$

$$\mathcal{L}_{1+}{}^{(3/2)B}=\frac{1}{k_0W}W^2O_{1+}|\mathbf{q}|\left(\frac{-g_rg_A(k^2)}{2M_N}\right)\left[\frac{M_N(p_{10}-M_N)W_++(\frac{1}{2}W_--p_{20})k^2}{W}\right.$$
$$\left.\times\frac{A(a)}{|\mathbf{q}|^2|\mathbf{k}|^2}+\frac{M_N(p_{10}+M_N)W_--(\frac{1}{2}W_+-p_{20})k^2}{(p_{10}+M_N)(p_{20}+M_N)W}\frac{C(a)}{|\mathbf{q}||\mathbf{k}|}\right],$$

$$\mathcal{H}_{1+(g_A)}{}^{(3/2)B}=O_{1+}|\mathbf{q}|\frac{g_rg_A(k^2)}{2M_N}\left[(\tfrac{1}{2}W_--q_0)\frac{A(a)}{|\mathbf{q}|^2|\mathbf{k}|}-\frac{(\frac{1}{2}W_+-q_0)}{(p_{10}+M_N)(p_{20}+M_N)}\frac{C(a)}{|\mathbf{q}|}\right], \tag{A1}$$

with

$$W_\pm=W\pm M_N,\quad O_{1+}=[(p_{10}+M_N)(p_{20}+M_N)]^{1/2},\quad O_{2+}=[(p_{10}+M_N)/(p_{20}+M_N)]^{1/2},$$
$$a=(2p_{20}k_0+k^2)/(2|\mathbf{q}||\mathbf{k}|),\quad \bar{a}=(2p_{20}q_0-M_\pi{}^2)/(2|\mathbf{q}|^2). \tag{A2}$$

The functions A through E are defined by

$$A(a)=1-\tfrac{1}{2}a\ln\left(\frac{a+1}{a-1}\right),\quad B(a)=\tfrac{1}{2}\left[a+\tfrac{1}{2}(1-a^2)\ln\left(\frac{a+1}{a-1}\right)\right],$$
$$C(a)=-\tfrac{1}{2}\left[3a+\tfrac{1}{2}(1-3a^2)\ln\left(\frac{a+1}{a-1}\right)\right],\quad E(a)=\tfrac{1}{2}\left[\tfrac{2}{3}-a^2+\tfrac{1}{2}a(a^2-1)\ln\left(\frac{a+1}{a-1}\right)\right], \tag{A3}$$

and $F_1{}^V(k^2)$ and $F_2{}^V(k^2)$ are, respectively, the isovector nucleon charge and magnetic form factors, normalized so that $F_1{}^V(0)+2M_NF_2{}^V(0)=4.7$. For reasons explained in Ref. 10, only the part of $M_{1+}{}^{(3/2)B}$ proportional to the total nucleon isovector magnetic moment (given explicitly in the equation above) is used in Eq. (25); the part proportional to the nucleon and pion charges should be dropped.

Reprinted from ANNALS OF PHYSICS
Vol. 50, No. 2, November 1968
Printed in Belgium

Photo-, Electro-, and Weak Single-Pion Production in the (3,3) Resonance Region

STEPHEN L. ADLER

Institute for Advanced Study, Princeton, New Jersey 08540

We give a unified account of single-pion photo-, electro-, and weak production. The emphases of the paper are fourfold: (1) We give a detailed kinematic discussion of single-pion electro- and weak-production; (2) we develop a dynamical model for electroproduction and weak production in the (3,3) resonance region, based on the CGLN model for photoproduction; (3) we systematically discuss the partially-conserved axail-vector current (PCAC) and current-algebra constraints which relate the single-pion electro- and weak-production matrix elements to the matrix elements for other processes; (4) we compare our model with experiment.

1. INTRODUCTION

Production of a single pion is the simplest inelastic process that can be studied in electron–nucleon and neutrino–nucleon scattering experiments. Already, several pion electroproduction experiments have been performed and some crude data on weak pion production is available. Since, in the future, there will be a substantial accumulation of data on these processes, the theoretical interpretation of pion production experiments becomes an important problem.

We give in this paper (*1*) a detailed, unified treatment of single pion photo-, electro- and weak production. The parallel discussion of the three processes is natural, since they are closely related. Photo-production is, of course, just a special case of electroproduction, in which a real photon, rather than a virtual photon, is involved. According to the conserved vector current (CVC) hypothesis (*2*), the isovector electroproduction amplitudes are related by an isospin rotation to the vector-current weak-production amplitudes. The weak production process also involves axial-vector amplitudes which, while not directly related to electroproduction amplitudes, are most naturally treated in analogy with the treatment of the vector amplitudes when making dynamical models. Comparison of photoproduction and electroproduction models with experiment gives an idea of how good weak-production models may be expected to be.

The main emphases of this paper are fourfold: First of all, we give a detailed *kinematic discussion* of single-pion electroproduction and weak production,

including a derivation of differential and total cross-section formulas for the weak-production case, and a discussion of the kinematic singularities which appear in the electroproduction matrix element when gauge invariance is imposed. The kinematic results are general, and are not limited to single-pion production in the (3, 3) resonance region. Secondly, we construct a *dynamical model* for the single-pion electro- and weak-production matrix elements in the (3, 3) resonance region. We limit our dynamical discussion to this region because, as shown in the pioneering work of Chew, Goldberger, Low, and Nambu (*3*) (CGLN), in the (3, 3) resonance region a very successful model for pion photoproduction can be made. Our work is essentially an extension to the electro- and weak-production cases of the version of the CGLN model discussed recently by Höhler and Schmidt (*4*). Thirdly, we discuss various *PCAC and current algebra constraints* which relate the single-pion electro- and weak-production matrix elements to the matrix elements for pion–nucleon scattering, for two-pion electroproduction and to the nucleon vector and axial-vector form factors. Wherever possible, we test whether our (3, 3)-dominated model satisfies the PCAC conditions. Finally, using predictions of our theoretical model in the (3, 3) resonance region, we give a *comparison with experimental data* obtained in recent photon, electron, and neutrino experiments.

The CGLN dispersion-theoretic dynamical model is not the only approach to getting predictions which we could have used. Other recent methods involve (i) use of a postulated higher symmetry, such as SU_6, to relate the N–$N_{3,3}^*$ (vector or axial-vector current) vertex to the nucleon vector and axial-vector form factors (*5*), or (ii) direct introduction of N–$N_{3,3}^*$ (vector or axial-vector current) form factors, which are parametrized in a convenient fashion and are used to generate a family of cross section curves (*6*). We prefer the CGLN approach because it has given more detailed and more accurate results than the higher symmetry method in the case of photoproduction, and because approach (ii) is little better than phenomenology unless specific dynamical assumptions, such as use of a higher symmetry or of the CGLN model, are made in order to give definite values for the N–$N_{3,3}^*$ (vector or axial-vector current) form factors. Of course, there exists already a considerable literature on the dispersive approach to single-pion electro- and weak production (*7*), (*8*). The most extensive recent treatments are the electroproduction calculation of Zagury (*7*), and a weak-production calculation by Salin (*8*) based on the work of Dennery (*7*), (*8*). In an Appendix we give a detailed comparison of our model with the work of Zagury and of Dennery. As noted above, we *only* discuss pion production in the (3, 3) resonance region. We do not treat higher isobar production, a topic which has been discussed recently by several authors (*9*).

Having explained the aims and scope of this paper, we turn next to a brief elaboration of its contents. Section 2 is devoted to a discussion of the kinematics of pion weak production and electroproduction; most of the subsection headings are self-explanatory. In order to keep this section readable, we have put most

detailed kinematic formulas in appendices. In Subsection 2D(3) we discuss only the most elementary consequences of the partially-conserved axial-vector current (PCAC) hypothesis, obtained by equating the divergence of the axial-vector weak pion production amplitude to the (off-shell) pion–nucleon scattering amplitude, and expressing the resulting equality in terms of multipoles. Other consequences of PCAC, including soft-pion theorems, are given in Section 5. Our principal new kinematic results are the formulas for the weak-pion-production differential cross section in terms of helicity amplitudes, and for the weak-production total cross section (integrated over the pion emission angles) expressed as a sum over multipole amplitudes, given in Subsection 2F and Appendix 4. In Subsection 2G and Appendix 5 we use the Rarita–Schwinger formalism to define direct N–$N_{3,3}^{*}$ (vector or axial-vector current) couplings, and we relate these couplings to the multipoles leading to the N to $N_{3,3}^{*}$ transition. This will enable the comparison of our paper, and other dispersive approach papers which calculate the multipole amplitudes leading to excitation of the (3, 3) resonance, with papers using the phenomenological, direct-coupling approach.

In Section 3 we write down the fixed momentum transfer dispersion relations which the weak production amplitudes obey and discuss the questions of kinematic singularities and convergence. We show that the use of gauge invariance to reduce the number of independent vector amplitudes from eight to six *necessarily* introduces a kinematic singularity in some of the invariant amplitudes, but that the effect of this singularity on the *physical matrix element* can be eliminated by an appropriate subtraction in the dispersion relations.

In Section 4 we develop a dynamical model for pion weak production and electroproduction. When specialized to pion photoproduction, our model differs only slightly from the Höhler–Schmidt version of the CGLN model. The general method is to write fixed momentum transfer dispersion relations for the invariant amplitudes. Under the dispersion integrals we approximate the imaginary parts of the amplitudes by keeping only multipoles which excite the (3, 3) resonance and which are dominant in Born approximation. We then project out integral equations for these multipoles; an examination of the nearest left-hand singularity structure of the multipoles shows enough of a resemblance to the familiar case of pion–nucleon scattering to allow a simple approximate solution to the integral equations. We guess this approximate solution by the heuristic procedure of first studying the static-nuleon limit of the integral equations. (At the same time we try to clarify the relation between several approaches found in the literature for solving the Omnes equations involved.) We then check numerically that the guess is a reasonably self-consistent solution to the integral equations when no static approximation is made, so that our final answer is *not* a static limit result. We show that the same arguments leading to our model for the dominant multipoles also can be used to justify an approximation which we have used elsewhere *(10)* for the pion off-shell

extrapolation of partial-wave and multipole amplitudes. Our model is summarized in Subsection 4E; the static limit of the model is calculated here only as a check, and is not used in the subsequent numerical work. A comparison of our model with other weak production and electroproduction calculations is given in Appendix 7.

In Section 5 we derive a large number of PCAC and current-algebra conditions on the pion electroproduction and weak production amplitudes, and compare them with values for the amplitudes calculated in our model. We interpret one particularly bad discrepancy as indicating that we have neglected a vector meson exchange contribution to weak pion production by the axial-vector current. We briefly discuss the low-energy theorem which relates two pion electroproduction, with one pion soft, to single-pion weak production and electroproduction amplitudes.

Finally, in Section 6 we compare our model with experiment. Agreement with photoproduction data and with electroproduction data for electron momentum transfers less than 0.6 $(\mathrm{BeV}/c)^2$ is good, but our theory appears to break down for momentum transfers much greater than 0.6 $(\mathrm{BeV}/c)^2$. In weak production, a fit of our model to CERN data for neutrino production of the (3, 3) resonance suggests an axial-vector form factor which falls off more slowly with increasing momentum transfer than do the vector form factors. We also discuss some features of weak pion production which may be of interest in future experiments.

2. KINEMATICS

In this section we discuss the kinematics of the pion weak, electro-, and photoproduction reactions. Many of the formulas give here have been published elsewhere; our aim has been to collect all of the kinematic equations needed in this work, using a standardized notation.

2A. Energy and Angle Variables

Let us consider the weak, electro- and photo- pion production reactions

$$\begin{aligned} \left.\begin{matrix}\nu_\ell \\ \bar{\nu}_\ell\end{matrix}\right\} (k_1) + N(p_1) &\to \left.\begin{matrix}\ell \\ \bar{\ell}\end{matrix}\right\} (k_2) + N(p_2) + \pi(q), \\ e(k_1) + N(p_1) &\to e'(k_2) + N(p_2) + \pi(q), \qquad (2A.1) \\ \gamma(k) + N(p_1) &\to N(p_2) + \pi(q). \end{aligned}$$

Substituting the static limit equations into Eq. (2F.8) for the differential cross section gives, after some algebraic rearrangement,

$$\frac{d^2\sigma}{d\Omega_\ell^{\rm L}\, dW} = \frac{k_{10}^{\rm L} k_{20}^{\rm L}}{\pi} \frac{d^2\sigma}{d(k^2)\, dW} \approx \frac{G_{\rm F}^2 \cos^2\theta_{\rm C}}{4\pi^3} \frac{(k_{10}-\omega)^2}{(\omega^2 - M_\pi^2)^{1/2}} \left(\frac{2M_N}{g_r}\right)^2 \sigma_{3,3}(W)\, \alpha, \tag{4E.6}$$

with

$$\sigma_{3,3}(W) = 16\pi(\tfrac{2}{3}a^{(+)} - \tfrac{1}{3}a^{(-)})^2\, |f_{1+}^{(3/2)}|^2 \tag{4E.7}$$

and with

$$\begin{aligned}\alpha = {} & g_A^2[1 + \sin^2(\theta/2)] + (F_1^V + 2M_N F_2^V)^2 \sin^2(\theta/2) \\ & \times [\tfrac{1}{2}(k_{10}^2 + k_{20}^2) + k_{10}k_{20}\sin^2(\theta/2)]/M_N^2 \\ & + 2\xi g_A(F_1^V + 2M_N F_2^V)\sin^2(\theta/2)(k_{10} + k_{20})/M_N - m_\ell^2 g_A h_A \omega/(2M_N k_{20}) \\ & + [h_A/(2M_N)]^2\, m_\ell^2[\omega^2 + 4k_{10}k_{20}\sin^2(\theta/2)][\sin^2(\theta/2) + m_\ell^2/(4k_{20}^2)].\end{aligned} \tag{4E.8}$$

As the notation suggests, $\sigma_{3,3}(W)$ has been defined so that in $(\nu/\bar{\nu})$-induced weak reactions it is the cross section for (3, 3) excitation in (π^+/π^-)-nucleon scattering. Eqs. (4E.6-8) are identical with the static model results of Bell and Berman. (Bell and Berman omit the lepton mass terms.)

5. WEAK PION PRODUCTION AND THE PARTIALLY-CONSERVED AXIAL-VECTOR CURRENT HYPOTHESIS

In this section we discuss in a systematic way the implications of the PCAC hypothesis for weak single pion production. We first derive the PCAC predictions, and then, wherever possible, compare them with the model for the pion production amplitude derived in the previous section. We interpret a glaring discrepancy between one of the PCAC predictions and our model as indicating that we omitted an important vector meson exchange contribution to weak pion production by the axial-vector current. Finally, we briefly discuss the connection between the reaction $e + N \rightarrow e' + N + \pi(\text{soft}) + \pi$ and single pion electro- and weak production.

5A. Derivation of the PCAC Relations

(1) *Small-k^2 Conditions (Axial-Vector Part)*

We saw in Subsection 2D that for small k^2 the PCAC hypothesis relates the axial-vector multipoles $\mathscr{L}_{\ell\pm}$ and $\mathscr{H}_{\ell\pm}$ to the corresponding pion–nucleon scattering partial-wave amplitudes $f_{\ell\pm}$ [see Eqs. (2D.12), (2D.15) and (2D.19)]. It will be useful to rewrite the identities contained in Eq. (2D.12) in terms of the invariant

and the center of mass amplitudes which were introduced in Subsection 2C. This is most easily done by going back to the statement of PCAC,

$$_{\text{out}}\langle \pi N \mid \partial_\lambda (J_\lambda^{A1} + iJ_\lambda^{A2}) \mid N\rangle = \frac{\sqrt{2}\, M_N M_\pi^{\,2} g_A}{g_r(0)}\ {}_{\text{out}}\langle \pi N \mid \varphi_\pi \mid N\rangle \tag{5A.1}$$

and expressing the right- and left-hand sides of this equation in terms of either invariant or center-of-mass frame amplitudes.

Invariant amplitudes:

Writing[33]

$$\begin{aligned}
&_{\text{out}}\langle \pi(q)\, N(p_2) \mid \varphi_\pi \mid N(p_1)\rangle \\
&\quad = \frac{1}{k^2 + M_\pi^{\,2}}\, \bar{u}_N(p_2)\{\sqrt{2}\, a^{(+)}[A^{\pi N(+)}(\nu, \nu_B, k^2, q^2 = -M_\pi^{\,2}) \\
&\qquad - i\gamma \cdot k B^{\pi N(+)}(\nu, \ldots)] + \sqrt{2}\, a^{(-)}[A^{\pi N(-)}(\nu, \ldots) - i\gamma \cdot k B^{\pi N(-)}(\nu, \ldots)]\}\, u_N(p_1)
\end{aligned} \tag{5A.2}$$

and expressing the left-hand side of Eq. (5A.1) in terms of the invariant amplitudes V_j and A_j, we find

$$\begin{aligned}
&-2M_N\nu(A_1 + A_2)^{(\pm)} + 2M_N\nu_B A_3^{(\pm)} + k^2 A_7^{(\pm)} \\
&\quad = \frac{2M_N g_A}{g_r(0)} \frac{M_\pi^{\,2}}{k^2 + M_\pi^{\,2}} A^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^{\,2}), \\
&2M_N A_1^{(\pm)} - M_N A_4^{(\pm)} + 2M_N\nu A_5^{(\pm)} - 2M_N\nu_B A_6^{(\pm)} - k^2 A_8^{(\pm)} \\
&\quad = \frac{2M_N g_A}{g_r(0)} \frac{M_\pi^{\,2}}{k^2 + M_\pi^{\,2}} B^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^{\,2}).
\end{aligned} \tag{5A.3}$$

The physical content of Eq. (5A.3) is, of course, just the same as that of Eq. (2D.12). Since we will want, in particular, to study Eq. (5A.3) at the point $\nu = \nu_B = 0$, we must separate off the Born terms, which become singular at that point. The pion–nucleon scattering Born approximation is

$$B^{\pi N(\pm)B}(\nu, \nu_B, k^2, -M_\pi^{\,2}) = \frac{g_r g_r(k^2)}{2M_N}\left(\frac{1}{\nu_B - \nu} \mp \frac{1}{\nu_B + \nu}\right) \tag{5A.4}$$

[33] Equation (5A.2) defines the off-shell pion-nucleon scattering amplitudes $A^{\pi N(\pm)}(\nu, \nu_B, k^2, q^2 = -M_\pi^{\,2})$ and $B^{\pi N(\pm)}(\nu, \nu_B, k^2, q^2 = -M_\pi^{\,2})$, in which the initial pion has $(\text{mass})^2 = -k^2$. They are related to the physical pion–nucleon scattering amplitudes $A^{\pi N(\pm)}(\nu, \nu_B)$ and $B^{\pi N(\pm)}(\nu, \nu_B)$ by

$$A^{\pi N(\pm)}(\nu, \nu_B, k^2 = -M_\pi^{\,2}, q^2 = -M_\pi^{\,2}) = A^{\pi N(\pm)}(\nu, \nu_B),$$
$$B^{\pi N(\pm)}(\nu, \nu_B, k^2 = -M_\pi^{\,2}, q^2 = -M_\pi^{\,2}) = B^{\pi N(\pm)}(\nu, \nu_B).$$

Our notation follows Adler (*10*).

and the weak production Born approximation is given in Eq. (2E.6). Using a bar to denote the non-Born part of the amplitude, e.g., $A_1^{(\pm)} = \bar{A}_1^{(\pm)} + A_1^{(\pm)B}$, we find by substituting Eqs. (5A.4) and (2E.6) into Eq. (5A.3) that[34]

$$\begin{aligned}
&-2M_N\nu(\bar{A}_1 + \bar{A}_2)^{(\pm)} + 2M_N\nu_B\bar{A}_3^{(\pm)} + k^2\bar{A}_7^{(\pm)} + \binom{1}{0}\, 2g_r g_A(k^2) \\
&\quad = \frac{2M_N g_A}{g_r(0)} \frac{M_\pi^2}{k^2 + M_\pi^2} \bar{A}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2), \\
&2M_N\bar{A}_1^{(\pm)} - M_N\bar{A}_4^{(\pm)} + 2M_N\nu\bar{A}_5^{(\pm)} - 2M_N\nu_B\bar{A}_6^{(\pm)} - k^2\bar{A}_8^{(\pm)} \\
&\quad = \frac{2M_N g_A}{g_r(0)} \frac{M_\pi^2}{k^2 + M_\pi^2} \bar{B}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2).
\end{aligned} \tag{5A.5}$$

Eq. (5A.5) can be further simplified by separating A_7 and A_8 into their one-pion-exchange contributions, coming from the diagram of Fig. 3, and a remainder,

$$\begin{aligned}
\bar{A}_7^{(\pm)} &= -\frac{2M_N g_A}{g_r(0)} \frac{1}{k^2 + M_\pi^2} \bar{A}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2) + \bar{A}_7^{(\pm)R}, \\
\bar{A}_8^{(\pm)} &= \frac{2M_N g_A}{g_r(0)} \frac{1}{k^2 + M_\pi^2} \bar{B}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2) + \bar{A}_8^{(\pm)R}.
\end{aligned} \tag{5A.6}$$

Equation (5A.5) becomes

$$\begin{aligned}
&-2M_N\nu(\bar{A}_1 + \bar{A}_2)^{(\pm)} + 2M_N\nu_B\bar{A}_3^{(\pm)} + k^2\bar{A}_7^{(\pm)R} + \binom{1}{0}\, 2g_r g_A(k^2) \\
&\quad = \frac{2M_N g_A}{g_r(0)} \bar{A}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2), \\
&2M_N\bar{A}_1^{(\pm)} - M_N\bar{A}_4^{(\pm)} + 2M_N\nu\bar{A}_5^{(\pm)} - 2M_N\nu_B\bar{A}_6^{(\pm)} - k^2\bar{A}_8^{(\pm)R} \\
&\quad = \frac{2M_N g_A}{g_r(0)} \bar{B}^{\pi N(\pm)}(\nu, \nu_B, k^2, -M_\pi^2).
\end{aligned} \tag{5A.7}$$

In Eq. (5A.7) the pion propagator $(k^2 + M_\pi^2)^{-1}$, which varies rapidly with k^2, has dropped out; the physical content of PCAC is the assertion that the left- and right-hand sides of Eq. (5A.7) vary only slowly as k^2 is varied in the interval $-M_\pi^2 \leqslant k^2 \leqslant 0$ (apart from possible threshold corrections of the type discussed in Subsection 2D).

From Eq. (5A.7) and the crossing properties of Eq. (2C.3), which state that certain of the amplitudes $\bar{A}_j$ vanish at $\nu = 0$, we deduce the following relations *(32)*:

$$\frac{g_r g_r(0)}{M_N} = \bar{A}^{\pi N(+)}\Big|_{\nu=\nu_B=k^2=0}; \tag{5A.8}$$

[34] We use the equation $2M_N g_A(k^2) - k^2 h_A(k^2) = [2M_N g_A/g_r(0)][M_\pi^2/(k^2 + M_\pi^2)]\, g_r(k^2)$, obtained by sandwiching Eq. (2D.10) between one nucleon states.

$$
\begin{aligned}
[\bar{A}_1^{(-)} - \tfrac{1}{2}\bar{A}_4^{(-)}]\Big|_{\nu=\nu_B=k^2=0} &= \frac{g_A}{g_r(0)} \bar{B}^{\pi N(-)}\Big|_{\nu=\nu_B=k^2=0}, \\
-[\bar{A}_1^{(-)} + \bar{A}_2^{(-)}]\Big|_{\nu=\nu_B=k^2=0} &= \frac{g_A}{g_r(0)} \frac{\partial \bar{A}^{\pi N(-)}}{\partial \nu}\Big|_{\nu=\nu_B=k^2=0}, \qquad (5A.9) \\
\bar{A}_3^{(+)}\Big|_{\nu=\nu_B=k^2=0} &= \frac{g_A}{g_r(0)} \frac{\partial \bar{A}^{\pi N(+)}}{\partial \nu_B}\Big|_{\nu=\nu_B=k^2=0}.
\end{aligned}
$$

Equation (5A.8) is the familiar consistency condition on πN scattering implied by PCAC (*33*). To interpret Eq. (5A.9) we note that, for small four-vector k, we have

$$
\begin{aligned}
&\bar{u}_N(p_2)\left[\sum_{j=1}^{6} O(V_j)\, V_j^{(\pm)} + \sum_{j=1}^{8} O(A_j)\, A_j^{(\pm)}\right] u_N(p_1) \\
&= \bar{u}_N(p_2)\Bigg\{\sum_{j=1}^{6} O(V_j)\, V_j^{(\pm)B} + \sum_{j=1}^{8} O(A_j)\, A_j^{(\pm)B} \\
&\quad + i[\bar{A}_1^{(\pm)} + \bar{A}_2^{(\pm)}]|_0\, (p_1 + p_2)\cdot e + i\bar{A}_3^{(\pm)}|_0\, q\cdot e \\
&\quad - [\bar{A}_4^{(\pm)} - 2\bar{A}_1^{(\pm)}]|_0\, M_N \gamma\cdot e + O(k)\Bigg\} u_N(p_1). \qquad (5A.10)
\end{aligned}
$$

The notation $|_0$ means that the invariant amplitudes are evaluated at $k = 0$, that is, at $\nu = \nu_B = k^2 = 0$. Since crossing symmetry implies that $\bar{A}_1^{(+)}|_0 = \bar{A}_2^{(+)}|_0 = \bar{A}_4^{(+)}|_0 = \bar{A}_3^{(-)}|_0 = 0$, the Born approximation and Eqs. (5A.9) completely specify the weak production amplitude, up to terms of first order in k.

Center-of-mass amplitudes:

Using the definition of the pion-nucleon scattering center of mass amplitudes f_1 and f_2,

$$
\bar{u}_N(p_2)[A^{\pi N} - i\gamma\cdot kB^{\pi N}]\, u_N(p_1) = \frac{4\pi W}{M_N} \chi_f^{*}[f_1 + \boldsymbol{\sigma}\cdot\hat{\mathbf{q}}\boldsymbol{\sigma}\cdot\hat{\mathbf{k}} f_2]\, \chi_i, \qquad (5A.11)
$$

we find that Eq. (5A.1) takes the form

$$
\begin{aligned}
\mathscr{G}_5^{A(\pm)} \frac{|\mathbf{k}|^2}{k_0} - \mathscr{G}_7^{A(\pm)} \frac{k^2}{k_0} &= \frac{8\pi W g_A}{g_r(0)} \frac{M_\pi^2}{k^2 + M_\pi^2} f_2^{(\pm)}, \\
\mathscr{G}_6^{A(\pm)} \frac{|\mathbf{k}|^2}{k_0} - \mathscr{G}_8^{A(\pm)} \frac{k^2}{k_0} &= \frac{8\pi W g_A}{g_r(0)} \frac{M_\pi^2}{k^2 + M_\pi^2} f_1^{(\pm)}.
\end{aligned} \qquad (5A.12)
$$

If f_1 and f_2 are expanded in partial waves according to

$$f_1 = \sum_{\ell=0}^{\infty} f_{\ell+} P'_{\ell+1}(y) - \sum_{\ell=2}^{\infty} f_{\ell-} P'_{\ell-1}(y).$$

$$f_2 = \sum_{\ell=1}^{\infty} (f_{\ell-} - f_{\ell+}) P'_{\ell}(y), \tag{5A.13}$$

comparison with the partial-wave expansion of $\mathscr{G}^A_{5,\ldots,8}$ in Eq. (2D.4) leads immediately to Eq. (2D.12).

(2) *Small-q Conditions (Axial-Vector and Vector Parts)*

As has been much discussed recently, the PCAC hypothesis, combined with the algebra of currents proposed by Gell-Mann (*34*), leads to a "pion low-energy theorem" for any weak or electromagnetic process in which a pion is emitted (*35*). The theorem relates the matrix element of the process, at zero-pion four-momentum, to the matrix element of the corresponding process in which no pion is present and to an equal time commutator of currents. As applied to weak or electroproduction of pions, the method gives restrictions on certain of the invariant amplitudes at the point $q = 0$. We give a detailed derivation of the restrictions on the axial-vector amplitudes A_j, and state (without giving a derivation) the similar results for the vector amplitudes V_j.

The low-energy theorem is derived from the identity

$$\begin{aligned} i \int d^4x e^{-iq\cdot x}(-\Box_x + M_\pi^2)\, \psi_c^* \langle N(p_2) \mid T[\partial_\sigma J_\sigma^{Ac}(x)(J_\lambda^{A1}(0) + iJ_\lambda^{A2}(0))] \mid N(p_1)\rangle\, e_\lambda \\ = -i(q^2 + M_\pi^2) \int d^3x e^{-iq\cdot x} \psi_c^* \\ \times \langle N(p_2) \mid [J_0^{Ac}(x), J_\lambda^{A1}\, 0) + iJ_\lambda^{A2}(0)]|_{x_0=0} \mid N(p_1)\rangle\, e_\lambda \\ - q_\sigma \int d^4x e^{-iq\cdot x}(-\Box_x + M_\pi^2)\, \psi_c^* \\ \times \langle N(p_2) \mid T[J_\sigma^{Ac}(x)(J_\lambda^{A1}(0) + iJ_\lambda^{A2}(0))] \mid N(p_1)\rangle\, e_\lambda\, , \end{aligned} \tag{5A.14}$$

obtained by integration by parts. We evaluate each of the terms in the limit as $q \to 0$. Using the PCAC hypothesis in the form

$$\partial_\sigma J_\sigma^{Ac}(x) = \frac{M_N M_\pi^2 g_A}{g_r(0)}\, \varphi_\pi{}^c(x), \tag{5A.15}$$

the left-hand side of Eq. (5A.14) is seen to be proportional to the matrix element for weak production of a pion of (mass)$^2 = -q^2$,

$$\text{left-hand side} = \frac{M_N M_\pi{}^2 g_A}{g_r(0)} \sum_{j=1}^{8} [a^{(+)} A_j^{(+)}(\nu, \nu_B, k^2, q^2) + a^{(-)} A_j^{(-)}(\nu, \nu_B, k^2, q^2)] \times \bar{u}_N(p_2)\, O(A_j)\, u_N(p_1). \tag{5A.16}$$

At $q = 0$ the invariants q^2, ν and ν_B are zero. Using the postulated commutation relation (34)

$$\left[\int d^3x J_4^{Aa}(x), J_\lambda^{Ab}(0)\right]\Big|_{x_0=0} = -\epsilon_{abc} J_\lambda^{Vc}(0), \tag{5A.17}$$

the first term on the right-hand side of Eq. (5A.14) becomes

$$-M_\pi{}^2 \int d^3x \psi_c^* \langle N(p_2) \mid [J_4^{Ac}(x), J_\lambda^{A1}(0) + iJ_\lambda^{A2}(0)]|_{x_0=0} \mid N(p_1)\rangle\, e_\lambda$$
$$= M_\pi{}^2 \bar{u}_N(p_2)[F_2{}^V(k^2)\, O(A_2) - M_N^{-1} \times (F_1{}^V(k^2) + 2M_N F_2{}^V(k^2))\, O(A_4)]\, u_N(p_1)\, a^{(-)}. \tag{5A.18}$$

Finally, in the limit as $q \to 0$ only the one-nucleon pole terms contribute to the term proportional to q_σ on the right-hand side of Eq. (5A.14). In other words, this term is

$$-M_\pi{}^2 \psi_c^* \bar{u}_N(p_2) \Big\{ ig_A \gamma \cdot q \gamma_5 \frac{\tau_c}{2} \frac{\gamma \cdot p_2 + iM_N}{-2p_2 \cdot q} [i\gamma_\lambda \gamma_5 g_A(k^2) + \gamma_5 k_\lambda h_A(k^2)]\, \tfrac{1}{2}(\tau_1 + i\tau_2)$$
$$+ \tfrac{1}{2}(\tau_1 + i\tau_2)[i\gamma_\lambda\gamma_5 g_A(k^2) + \gamma_5 k_\lambda h_A(k^2)] \times \frac{\gamma \cdot p_1 + iM_N}{2p_1 \cdot q}\, ig_A \gamma \cdot q \gamma_5 \frac{\tau_c}{2} \Big\}\, u_N(p_1)\, e_\lambda + O(q). \tag{5A.19}$$

After some algebra, this can be rewritten in the form

$$M_\pi{}^2 g_A \bar{u}_N(p_2)[M_N^{-1} g_A(k^2)\, a^{(-)} O(A_4) - h_A(k^2)\, a^{(+)} O(A_7)]\, u_N(p_1)$$
$$+ \frac{M_N M_\pi{}^2 g_A}{g_r(0)} \bar{u}_N(p_2) \Big\{ \frac{g_r(0)\, g_A(k^2)}{2M_N} O(A_1) \times \left[a^{(+)}\left(\frac{1}{\nu_B - \nu} - \frac{1}{\nu_B + \nu}\right) + a^{(-)}\left(\frac{1}{\nu_B - \nu} + \frac{1}{\nu_B + \nu}\right)\right]$$
$$+ \frac{g_r(0)\, g_A(k^2)}{2M_N} O(A_3) \left[a^{(+)}\left(\frac{1}{\nu_B - \nu} + \frac{1}{\nu_B + \nu}\right) + a^{(-)}\left(\frac{1}{\nu_B - \nu} - \frac{1}{\nu_B + \nu}\right)\right]$$
$$+ \frac{g_r(0)\, h_A(k^2)}{2M_N} O(A_8) \times \left[a^{(+)}\left(\frac{1}{\nu_B - \nu} - \frac{1}{\nu_B + \nu}\right) + a^{(-)}\left(\frac{1}{\nu_B - \nu} + \frac{1}{\nu_B + \nu}\right)\right]\Big\}\, u_N(p_1), \tag{5A.20}$$

with the term in curly brackets just the Born approximation for weak pion production. Substituting Eqs. (5A.16), (5A.18), and (5A.20) into the identity of Eq. (5A.14), we see that the Born approximation terms, which are singular at $q = 0$, cancel out. This leaves us with the following conditions on the non-Born parts of the amplitudes (denoted again by a bar) *(36)*:

$$\bar{A}_2^{(-)}(\nu = \nu_B = 0, k^2, q^2 = 0) = \frac{g_r(0)}{M_N g_A} F_2{}^V(k^2),$$

$$\bar{A}_4^{(-)}(\nu = \nu_B = 0, k^2, q^2 = 0) = -\frac{g_r(0)}{M_N{}^2 g_A}[F_1{}^V(k^2) - g_A g_A(k^2) + 2M_N F_2{}^V(k^2)],$$

$$\bar{A}_7^{(+)}(\nu = \nu_B = 0, k^2, q^2 = 0) = -\frac{g_r(0)}{M_N} h_A(k^2). \tag{5A.21}$$

An entirely analogous derivation can be carried out for the vector weak production (and the electroproduction) amplitudes, leading to the identities *(37)*

$$\begin{aligned}
\bar{V}_1^{(+)}(\nu = \nu_B = 0, k^2, q^2 = 0) &= \frac{g_r(0)}{M_N} F_2{}^V(k^2),\\
\bar{V}_1^{(0)}(\nu = \nu_B = 0, k^2, q^2 = 0) &= \frac{g_r(0)}{M_N} F_2{}^S(k^2),\\
\bar{V}_6^{(-)}(\nu = \nu_B = 0, k^2, q^2 = 0) &= \frac{g_r(0)}{M_N}\left[\frac{g_A(k^2)}{g_A(0)} - F_1{}^V(k^2)\right](k^2)^{-1}.
\end{aligned} \tag{5A.22}$$

Equations (5A.21) and (5A.22), together with the Born approximation, completely specify the weak production amplitude up to terms of first order in q.

The condition on $\bar{A}_7^{(+)}$ has a simple interpretation when k^2 is near $-M_\pi{}^2$, so that only the pion pole terms in $\bar{A}_7^{(+)}$ and h_A need be retained. In the pion pole approximation,

$$\begin{aligned}
&\bar{A}_7^{(+)}(\nu = \nu_B = 0, k^2, q^2 = 0)\\
&\quad\approx -\frac{2M_N g_A}{g_r(0)}\frac{1}{k^2 + M_\pi{}^2}\bar{A}^{\pi N(+)}(\nu = \nu_B = 0, k^2 = -M_\pi{}^2, q^2 = 0),\\
&\qquad h_A(k^2) \approx \frac{2M_N g_A[g_r/g_r(0)]}{k^2 + M_\pi{}^2},
\end{aligned} \tag{5A.23}$$

and substituting these relations into Eq. (5A.21) gives

$$\bar{A}^{\pi N(+)}(\nu = \nu_B = 0, k^2 = -M_\pi{}^2, q^2 = 0) = \frac{g_r g_r(0)}{M_N}. \tag{5A.24}$$

Eq. (5A.24) is identical with the pion–nucleon scattering consistency condition of Eq. (5A.8). (It makes no difference whether it is the initial or the final pion which is off mass shell.)

(3) *Combined Relations*

A number of interesting relations can be obtained by combining the small-q equations of Eq. (5A.21) with the small-k equations of Eq. (5A.9). [More properly, we use the analog of Eq. (5A.9) in which the final pion mass $-q^2$ has been extrapolated from M_π^2 to 0. We neglect a small additional term, the so-called "σ term," which appears in the equation for $\bar{A}_3^{(+)}$ when $q^2 \neq -M_\pi^2$ *(38)*.] Taking the linear combination of $\bar{A}_2^{(-)}$ and $\bar{A}_4^{(-)}$ which eliminates $F_2^V(k^2)$ gives

$$2\bar{A}_2^{(-)}|_0 + \bar{A}_4^{(-)}|_0 = -g_r(0)(1 - g_A^2)/(M_N^2 g_A), \tag{5A.25}$$

while eliminating $\bar{A}_1^{(-)}$ from Eq. (5A.9) gives

$$-\bar{A}_2^{(-)}\Big|_0 - \frac{1}{2}\bar{A}_4^{(-)}\Big|_0 = \left[\frac{g_A}{g_r(0)}\right]\left[\frac{\partial \bar{A}^{\pi N(-)}}{\partial \nu} + \bar{B}^{\pi N(-)}\right]\Big|_0. \tag{5A.26}$$

[The abbreviation $|_0$ indicates evaluation of the amplitudes at $\nu = \nu_B = k^2 = q^2 = 0$.] Taken together, Eqs. (5A.25) and (5A.26) imply that

$$1 - \frac{1}{g_A^2} = -\frac{2M_N^2}{g_r^2(0)}\left[\frac{\partial \bar{A}^{\pi N(-)}}{\partial \nu} + \bar{B}^{\pi N(-)}\right]\Big|_{\nu=\nu_B=k^2=q^2=0}, \tag{5A.27}$$

the usual g_A sum rule *(39)*. In other words, the g_A sum rule emerges as the condition that the small-q and small-k expressions for the axial-vector weak production matrix element be consistent at the point $q = k = 0$ *(32)*.

Since Eqs. (5A.9), (5A.21) and (5A.22) determine the terms in the weak production matrix element linear in *either* q or k, one can use them to write down an expression for the weak production matrix element, exact up to terms of order qk, q^2 and k^2. Dropping lepton mass corrections, one finds *(32)*

$$\bar{u}_N(p_2)\left[\sum_{j=1}^{6} O(V_j)\, V_j^{(\pm)} + \sum_{j=1}^{6} O(A_j)\, A_j^{(\pm)}\right] u_N(p_1)$$

$$= \bar{u}_N(p_2)\left[\sum_{j=1}^{6} O(V_j)\, V_j^{(\pm)B} + \sum_{j=1}^{6} O(A_j)\, A_j^{(\pm)B} + \Delta^{(\pm)} + O(qk, q^2, k^2)\right] u_N(p_1),$$

$$\Delta^{(+)} = \frac{ig_A}{g_r(0)}\frac{\partial \bar{A}^{\pi N(+)}}{\partial \nu_B}\Big|_0 q\cdot e - \frac{g_r(0)\mu^V}{2M_N^2}\gamma_5 e_\alpha \sigma_{\alpha\beta} k_\beta, \tag{5A.28}$$

$$\Delta^{(-)} = \frac{ig_A}{g_r(0)}\left\{\frac{g_r(0)^2}{2M_N^2}\left(1 - \frac{1}{g_A^2}\right)(p_1 + p_2)\cdot e\right.$$

$$\left. - ie_\alpha\sigma_{\alpha\beta}q_\beta \bar{B}^{\pi N(-)}|_0 + \frac{g_r(0)^2}{2M_N^2} ie_\alpha\sigma_{\alpha\beta}k^\beta\left(1 - \frac{1}{g_A^2} - \frac{\mu^V}{g_A^2}\right)\right\}.$$

This result, while valid near $q = k = 0$ (and perhaps even good at the pion production threshold) is sure to fail in the (3, 3) resonance region, since it does not take final-state interactions into account. Thus, the small-q, -k expansions will not be of practical use in calculating weak pion production cross sections.

Another useful relation obtained from Eqs. (5A.9) and (5A.21) is (*40*)

$$\bar{A}_1^{(-)}\Big|_0 = -\frac{g_A}{g_r(0)}\frac{\partial\bar{A}^{\pi N(-)}}{\partial\nu}\Big|_0 - \frac{g_r(0)}{M_N g_A}F_2^V(0), \tag{5A.29}$$

or equivalently, by use of Eq. (5A.27)

$$\bar{A}_1^{(-)}\Big|_0 = \frac{g_A}{g_r(0)}\bar{B}^{\pi N(-)}\Big|_0 - \frac{g_r(0)}{2M_N^2}\Big[\frac{1}{g_A} - g_A + \frac{2M_N F_2^V(0)}{g_A}\Big]. \tag{5A.30}$$

If the weak production amplitude $\bar{A}_1^{(-)}|_0$ were zero or negligibly small, Eq. (5A.29) or Eq. (5A.30) would give a relation between the isovector nucleon magnetic moment and pion–nucleon scattering.[35] However, the numerical analysis of the next subsection shows that our weak production model does not give any theoretical reason for neglecting $\bar{A}_1^{(-)}|_0$.

5B. Comparison with Weak Production Model

In Table V we compare the PCAC predictions for the various covariant amplitudes with the values calculated from the model given in the previous section, at the point $\nu = \nu_B = k^2 = q^2 = 0$. The amplitudes $\bar{V}_1^{(+)}$, $\bar{A}_1^{(-)}$,... in Column (A) are calculated directly from Eq. (3A.2). The bar, we recall, means that only the non-Born part of the amplitude is retained. In our model, the non-Born part of the amplitude comes from the dispersion integrals over the dominant (3, 3) multipoles, which in turn are given by Eq. (4D.22). Since we actually need the dominant multipoles at the off-mass-shell point $q^2 = 0$, we use the off-mass-shell form of Eq. (4D.22),

$$M_{1+(u)}^{(3/2)}|_{q^2=0} = M_{1+(u)}^{(3/2)B}|_{q^2=0}[f_{1+}^{(3/2)}/f_{1+}^{(3/2)B}][1 + a(k^2)^2/(\omega\omega_{3,3})] \tag{5B.1}$$

and similarly for $\mathscr{E}_{1+}^{(3/2)}|_{q^2=0}$ and $\mathscr{L}_{1+}^{(3/2)}|_{q^2=0}$. The factor $g_r(0)^{-1}$ multiplying all the amplitudes in Column (A) of Table V cancels the factor $g_r(0)$ in $M_{1+(\mu)}^{(3/2)B}|_{q^2=0}$, $\mathscr{E}_{1+}^{(3/2)B}|_{q^2=0}$ and $\mathscr{L}_{1+}^{(3/2)B}|_{q^2=0}$, and thus drops out of the numerical evaluation.

[35] A number of authors (*41*), because of incorrectly using amplitudes with kinematic singularities, have obtained Eqs. (5A.29) and (5A.30) with $\bar{A}_1^{(-)}|_0$ replaced by 0.

The PCAC predictions in Column (B) were obtained from the experimental values

$$\begin{aligned} \mu^V &= 3.70, \\ \mu^S &= -0.12, \\ g_A &= 1.18, \\ F_1^{V\prime}(0) &= -0.045/M_\pi^2 \end{aligned} \tag{5B.2}$$

and from pion–nucleon scattering phase-shift data.[36]

For some of the amplitudes, the agreement between the low-energy predictions of PCAC and our weak production model is poor, indicating that, at least in the region near $\nu = \nu_B = k^2 = q^2 = 0$, significant omissions have been made from the weak production amplitude. Let us consider the entries in Table V individually:

$\zeta \bar{V}_1^{(+)}|_0$: The prediction of the model, 0.38, is 70% of the value 0.55 predicted by PCAC. A detailed analysis of the PCAC prediction for $\bar{V}_1^{(+)}|_0$ has been made by Adler and Gilman (*37*), who find that, in addition to the multipole $M_{1+}^{(3/2)}$, the multipoles $E_{1+}^{(3/2)}$, $E_{0+}^{(3/2)}$, and $E_{0+}^{(1/2)}$ also make significant contributions to the dispersion integral for $\bar{V}_1^{(+)}|_0$. Using experimental values for the important multipoles in the regions of the (3, 3) resonance and the second pion–nucleon resonance [$N^*(1520)$], Adler and Gilman found $\zeta \bar{V}_1^{(+)}|_0 = 0.47$.

$\zeta \bar{V}_1^{(0)}|_0$: In our model the isoscalar amplitude is pure Born approximation, so the barred, or non-Born, amplitude vanishes.

$\zeta \bar{V}_6^{(-)}|_0$: The analysis of Adler and Gilman (*37*) shows that the contribution of the $M_{1+}^{(3/2)}$ multipole to the dispersion relation for $\bar{V}_6^{(-)}|_0$ is kinematically suppressed, and consequently is smaller numerically than the contributions of the $E_{1+}^{(3/2)}$, $L_{1+}^{(3/2)}$ and other multipoles. This means that use of the magnetic dipole

[36] The quoted value of $\zeta^2 \bar{B}^{\pi N(-)}|_0$ was obtained from the analysis of Höhler and Strauss (*42*), who make the approximation

$$\zeta^2 \bar{B}^{\pi N(-)}|_0 \approx \bar{B}^{\pi N(-)}(0, -M_\pi^2/2M_N, -M_\pi^2, -M_\pi^2)$$

and calculate the physical amplitude on the right-hand side from phase shift data and a Regge model for the high-energy region. Similarly, to calculate $\zeta^2\, \partial \bar{A}^{\pi N(+)}/\partial \nu_B|_0$, we make the approximation

$$\begin{aligned} \zeta^2\, \partial \bar{A}^{\pi N(+)}/\partial \nu_B|_0 &\approx \partial \bar{A}^{\pi N(+)}(0, \nu_B, -M_\pi^2, -M_\pi^2)/\partial \nu_B|_{\nu_B = -M_\pi^2/2M_N} \\ &\approx \frac{2M_N}{M_\pi^2}[\bar{A}^{\pi N(+)}(0, 0, -M_\pi^2, -M_\pi^2) - \bar{A}^{\pi N(+)}(0, -M_\pi^2/2M_N, -M_\pi^2, -M_\pi^2)] \end{aligned}$$

and use the value

$$\bar{A}^{\pi N(+)}(0, 0, -M_\pi^2, -M_\pi^2) - \bar{A}^{\pi N(+)}(0, -M_\pi^2/2M_N, -M_\pi^2, -M_\pi^2) \approx 2.66/M_\pi$$

calculated from phase shift data by Adler (*43*).

TABLE V

COMPARISON OF PCAC PREDICTIONS WITH THE MODEL DEVELOPED IN SECTION 4[a]

(A) Value in Model $[\zeta = g_r/g_r(0)]$	(B) PCAC Prediction	Equation No.
$\zeta \bar{V}_1^{(+)}\vert_0 = 0.38$	$\frac{g_r}{M_N} F_2^V(0) = \frac{g_r \mu^V}{2M_N^2} = 0.55$	5A.22
$\zeta \bar{V}_1^{(0)}\vert_0 = 0$	$\frac{g_r}{M_N} F_2^S(0) = \frac{g_r \mu^S}{2M_N^2} = -0.018$	5A.22
$\zeta \bar{V}_6^{(-)}\vert_0 = 0.012$	$\frac{g_r}{M_N}\left[\frac{g_A'(0)}{g_A(0)} - F_1^{V\prime}(0)\right] = 0.090 - \frac{4}{M_A^2} = \begin{cases} 0.012 & \text{for } M_A = 1 \text{ BeV;} \\ 0.071 & \text{for } M_A = 2 \text{ BeV.}^b \end{cases}$	5A.22
$\zeta \bar{A}_1^{(-)}\vert_0 = 0.32$	$\frac{g_A}{g_r}\left[\zeta^2 \bar{B}^{\pi N(-)}\vert_0 - \frac{g_r^2}{2M_N^2 g_A^2}(1 - g_A^2 + \mu^V)\right] = 0.28^c$	5A.30
$\zeta \bar{A}_2^{(-)}\vert_0 = 0.12$	$\frac{g_r \mu^V}{2M_N^2 g_A} = 0.47$	5A.21
$\zeta \bar{A}_3^{(+)}\vert_0 = 1.3$	$\frac{g_A}{g_r} \zeta^2 \frac{\partial \bar{A}^{\pi N(+)}}{\partial \nu_B}\Big\vert_0 = 3.1^d$	5A.9
$\zeta \bar{A}_4^{(-)}\vert_0 = -0.096$	$\frac{-g_r}{M_N^2 g_A}[1 - g_A^2 + \mu^V] = -0.83$	5A.21
$\zeta \bar{A}_7^{(+)}\vert_0 / h_A(0) = -1.7$	$\frac{-g_r}{M_N} = -2.0$	5A.21

[a] $M_\pi = 1$ throughout.
[b] We have parametrized $g_A(k^2)$ in the form $g_A(k^2)/g_A(0) = (1 + k^2/M_A^2)^{-2}$.
[c] Obtained from the analysis of Höhler and Strauss (*42*). See Footnote 36.
[d] Obtained from the analysis of Adler (*43*). See Footnote 36.

dominance approximation in the dispersion relation for $V_6^{(-)}$ is dubious, and that comparison of the magnetic dipole result $\zeta \bar{V}_6^{(-)}\vert_0 = 0.012$ with the PCAC prediction has little meaning.

$\zeta \bar{A}_1^{(-)}\vert_0$: The PCAC prediction here is in reasonable agreement with the value given by the model. Since the value of $\zeta \bar{A}_1^{(-)}\vert_0$ in the model (0.32) is of the same order of magnitude as the magnetic moment term in the PCAC prediction (−0.47), the model gives no theoretical reason for the neglect of the weak pion production terms in Eq. (5A.30) relative to the magnetic moment and the pion-nucleon scattering terms.

$\zeta \bar{A}_2^{(-)}|_0$ and $\zeta \bar{A}_4^{(-)}|_0$: For each of these amplitudes individually, the model disagrees badly with the PCAC prediction. However, for the linear combination $\zeta[2\bar{A}_2^{(-)}|_0 + \bar{A}_4^{(-)}|_0]$ which enters into the g_A sum rule [see Eqs. (5A.25-27)], the prediction of the model is 0.14, in good agreement with the PCAC prediction of $-g_r(1 - g_A{}^2)/(M_N{}^2 g_A) = 0.10$.

$\zeta \bar{A}_3^{(+)}|_0$: The prediction of the model here is in fair agreement with PCAC.

$\zeta \bar{A}_7^{(+)}|_0/h_A(0)$: Here the integral over the (3, 3) resonance[37] is in good agreement with PCAC. However, this is somewhat of an accident, since as we noted in Subsection 3C, $A_7^{(+)}(\nu, \nu_B, k^2)$ does not satisfy an unsubtracted dispersion relation in ν! The significance of the good agreement is that the two terms in the subtraction constant of Eq. (3C.3) nearly cancel when $\nu_B = k^2 = 0$, making the subtraction constant small.

The comparison of our model with the PCAC predictions indicates that while agreement in the case of the photoproduction amplitudes $\bar{V}_1^{(+)}|_0$ and $\bar{V}_1^{(0)}|_0$ is good, agreement for most of the weak production amplitudes is less than satisfactory. Thus, it may *not* be correct to justify our model for weak production by its success in photoproduction, since the comparison with the PCAC predictions indicates that in the weak production case, important pieces of the amplitude have been omitted. However, this problem may not be as serious as it appears from Table V. The worst discrepancies occur in the amplitudes $\bar{A}_2^{(-)}$ and $\bar{A}_4^{(-)}$; we will show below that the trouble with these amplitudes comes from neglecting certain vector meson exchange contributions to weak pion production. While the vector exchange terms make the major contribution to $\bar{A}_2^{(-)}|_0$ and $\bar{A}_4^{(-)}|_0$, we shall see in Subsection 6C that they do not greatly change the weak pion production cross sections in the (3, 3) region.

5C. Vector Meson Exchange Amplitude

In this Subsection we calculate the vector meson exchange contribution to weak pion production by the axial-vector current (*44*). We will not limit ourselves to ρ exchange alone, but rather will sum over all diagrams in which a particle with the quantum numbers $J^{PG} = 1^{-+}$ is exchanged. As is discussed above in Subsection 3C, such t-channel singularities are not in general correctly included when the s-channel dispersion integrals (the integrals over x') are extended only over the

[37] The equations relating Im A_7 to Im $\mathscr{H}_{1+}^{\pi(3/2)}$, analogous to Eqs. (4B.6-8), are

$$\operatorname{Im} A_7^{(\pm)}[x, \nu_B, k^2] = \begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix} a_7,$$

$$a_7 = -2M_N W[W_-(p_{10} + M_N)(p_{20} + M_N) + 3W_+(2M_N\nu_B + q_0 k_0)] \times \operatorname{Im} \mathscr{H}_{1+}^{\pi(3/2)}/(W^2 O_{1+} |\mathbf{q}| |\mathbf{k}| k_0).$$

PHYSICAL REVIEW VOLUME 177, NUMBER 5 25 JANUARY 1969

Axial-Vector Vertex in Spinor Electrodynamics

STEPHEN L. ADLER
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 24 September 1968)

Working within the framework of perturbation theory, we show that the axial-vector vertex in spinor electrodynamics has anomalous properties which disagree with those found by the formal manipulation of field equations. Specifically, because of the presence of closed-loop "triangle diagrams," the divergence of axial-vector current is not the usual expression calculated from the field equations, and the axial-vector current does not satisfy the usual Ward identity. One consequence is that, even after the external-line wave-function renormalizations are made, the axial-vector vertex is still divergent in fourth- (and higher-) order perturbation theory. A corollary is that the radiative corrections to $\nu_l l$ elastic scattering in the local current-current theory diverge in fourth (and higher) order. A second consequence is that, in massless electrodynamics, despite the fact that the theory is invariant under γ_5 transformations, the axial-vector current is not conserved. In an Appendix we demonstrate the uniqueness of the triangle diagrams, and discuss a possible connection between our results and the $\pi^0 \to 2\gamma$ and $\eta \to 2\gamma$ decays. In particular, we argue that as a result of triangle diagrams, the equations expressing partial conservation of axial-vector current (PCAC) for the neutral members of the axial-vector-current octet must be modified in a well-defined manner, which completely alters the PCAC predictions for the π^0 and the η two-photon decays.

INTRODUCTION

THE axial-vector vertex in spinor electrodynamics is of interest because of its connections (i) with radiative corrections to $\nu_l l$ scattering and (ii) with the γ_5 invariance of massless electrodynamics. We will show in this paper, within the framework of perturbation theory, that the axial-vector vertex has anomalous properties which disagree with those found by the formal manipulation of field equations. In particular, because of the presence of closed-loop "triangle diagrams," the divergence of the axial-vector current is not the usual expression calculated from the field equations, and the axial-vector current does not satisfy the usual Ward identity. One consequence is that, even after external-line wave-function renormalizations are made, the axial-vector vertex is still divergent in fourth- (and higher-) order perturbation theory. A corollary is that the radiative corrections to $\nu_l l$ elastic scattering in the local current-current theory diverge in fourth (and higher) order. A second consequence is that, in massless electrodynamics, despite the fact that the theory is invariant under γ_5 transformations, the axial-vector current is not conserved.

In Sec. I we derive the usual formulas for the axial-vector divergence and Ward identity, and then show how they are modified by the presence of triangle diagrams. In Sec. II we discuss various consequences of the additional term found in Sec. I. In the Appendix we show that it is *not* possible to redefine the triangle diagram in a physically acceptable way so as to eliminate the anomalous behavior discussed in Secs. I and II. We also discuss in the Appendix a possible connection between our results and the $\pi^0 \to 2\gamma$ and $\eta \to 2\gamma$ decays. In particular, we argue that as a result of triangle diagrams, the equations expressing partial conservation of axial-vector current (PCAC) for the neutral members of the axial-vector current octet must be modified in a well-defined manner, which completely alters the PCAC predictions for the π^0 and the η two-photon decays.

I. AXIAL CURRENT DIVERGENCE AND WARD IDENTITY

We work in the usual spinor electrodynamics, described by the Lagrangian density[1]

$$\mathcal{L}(x)=\bar{\psi}(x)(i\gamma\cdot\square-m_0)\psi(x)-\tfrac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) -:e_0\bar{\psi}(x)\gamma_\mu\psi(x)A^\mu(x):, \quad (1)$$

$$F_{\mu\nu}(x)\equiv\frac{\partial A_\mu(x)}{\partial x^\nu}-\frac{\partial A_\nu(x)}{\partial x^\mu}, \quad \gamma\cdot\square\equiv\gamma^\mu\frac{\partial}{\partial x^\mu}.$$

We define the axial-vector current $j_\mu{}^5(x)$ and the pseudoscalar density $j^5(x)$ by

$$\begin{aligned} j_\mu{}^5(x)&=:\bar{\psi}(x)\gamma_\mu\gamma_5\psi(x):,\\ j^5(x)&=:\bar{\psi}(x)\gamma_5\psi(x):; \end{aligned} \quad (2)$$

the corresponding vertex parts $\Gamma_\mu{}^5(p,p')$ and $\Gamma^5(p,p')$ are defined by

$$\begin{aligned} &S_F'(p)\Gamma_\mu{}^5(p,p')S_F'(p')\\ &\quad=-\int d^4x d^4y\, e^{ip\cdot x}e^{-ip'\cdot y}\langle T(\psi(x)j_\mu{}^5(0)\bar{\psi}(y))\rangle_0,\\ &S_F'(p)\Gamma^5(p,p')S_F'(p')\\ &\quad=-\int d^4x d^4y\, e^{ip\cdot x}e^{-ip'\cdot y}\langle T(\psi(x)j^5(0)\bar{\psi}(y))\rangle_0. \end{aligned} \quad (3)$$

Using the equations of motion which follow from Eq. (1), the divergence of the axial-vector current may

[1] We use the notation and metric conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Co., New York, 1965), pp. 377–390. Note that $\epsilon_{0123}=-\epsilon^{0123}=1$.

easily be calculated to be

$$\frac{\partial}{\partial x_\mu} j_\mu^5(x) = 2im_0 j^5(x)\,. \tag{4}$$

From Eqs. (3) and (4), we obtain the usual axial-vector Ward identity

$$(p-p')^\mu \Gamma_\mu^5(p,p') = 2m_0 \Gamma^5(p,p') + S_F'(p)^{-1}\gamma_5 + \gamma_5 S_F'(p')^{-1}. \tag{5}$$

Our task in this section is to see whether Eqs. (4) and (5), which we have formally derived from the field equations, actually hold in perturbation theory.

To this end, let us rederive Eq. (5) in perturbation theory. It is convenient to write

$$\begin{aligned} \Gamma_\mu^5 &= \gamma_\mu\gamma_5 + \Lambda_\mu^5\,,\\ \Gamma^5 &= \gamma_5 + \Lambda^5\,,\\ S_F'(p)^{-1} &= p - m_0 - \Sigma(p)\,, \end{aligned} \tag{6}$$

where the vertex corrections Λ_μ^5 and Λ^5 and the proper self-energy part $\Sigma(p)$ are calculated using $(p-m_0)^{-1}$ as the free propagator. (Use of the bare mass $m_0 = m - \delta m$ in the free propagator automatically includes the mass-renormalization counter terms.) In terms of Λ_μ^5, Λ^5, and Σ, Eq. (5) becomes

$$(p-p')^\mu \Lambda_\mu^5(p,p') = 2m_0\Lambda^5(p,p') - \Sigma(p)\gamma_5 - \gamma_5\Sigma(p')\,. \tag{7}$$

In order to derive Eq. (7), let us divide the diagrams contributing to $\Lambda_\mu^5(p,p')$ into two types: (a) diagrams in which the axial-vector vertex $\gamma_\mu\gamma_5$ is attached to the fermion line beginning with external four-momentum p' and ending with external four-momentum p; (b) diagrams in which the axial-vector vertex $\gamma_\mu\gamma_5$ is attached to an internal closed loop [See Figs. 1(a) and 1(b), respectively]. A typical contribution of type (a) has the form

$$\sum_{k=1}^{2n-1}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(k)}\frac{1}{p+p_k-m_0}\gamma_\mu\gamma_5\frac{1}{p'+p_k-m_0} \times\prod_{j=k+1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}(\cdots)\,, \tag{8}$$

where we have focused our attention on the line to which the $\gamma_\mu\gamma_5$ vertex is attached and have denoted the remainder of the diagram by $(\cdots)$. Multiplying Eq. (8) by $(p-p')^\mu$ and making use of the identity

$$\frac{1}{p+p_k-m_0}(p-p')\gamma_5\frac{1}{p'+p_k-m_0} = \frac{1}{p+p_k-m_0}(2m_0\gamma_5) \times\frac{1}{p'+p_k-m_0} + \frac{1}{p+p_k-m_0}\gamma_5 + \gamma_5\frac{1}{p'+p_k-m_0} \tag{9}$$

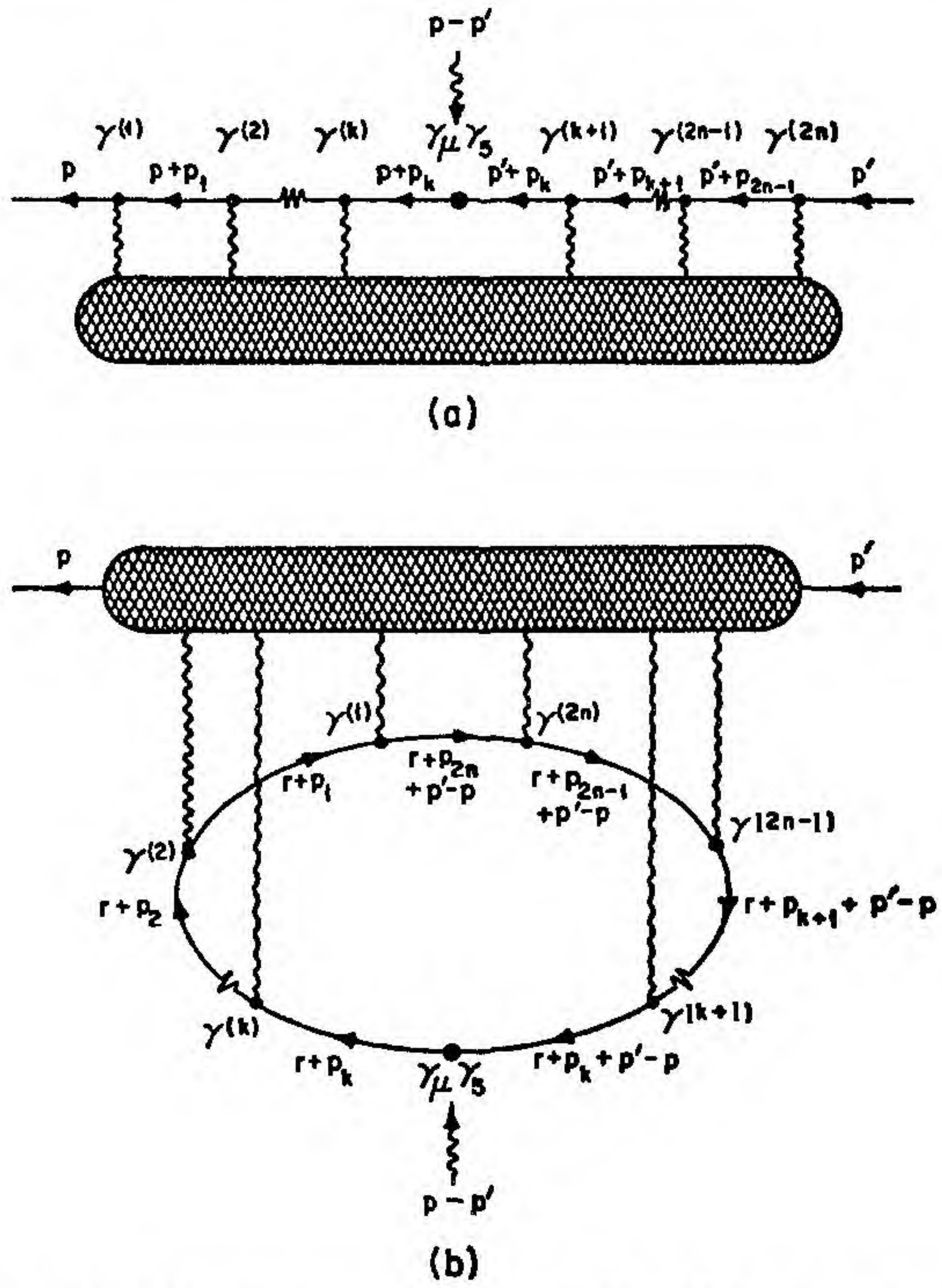


FIG. 1. Diagrams contributing to the axial-vector vertex. (a) The axial-vector vertex is attached to the fermion line beginning with external four-momentum p' and ending with external four-momentum p. (b) The axial-vector vertex is attached to an internal closed loop.

gives, after a little algebraic rearrangement,

$$\begin{aligned} &\sum_{k=1}^{2n-1}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(k)}\frac{1}{p+p_k-m_0}2m_0\gamma_5 \\ &\quad\times\frac{1}{p'+p_k-m_0}\prod_{j=k+1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}(\cdots) \\ &\quad-(\cdots)\prod_{j=1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(2n)}\gamma_5 \\ &\quad-\gamma_5\prod_{j=1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}(\cdots)\,. \end{aligned} \tag{10}$$

The first, second, and third terms in Eq. (10) are, respectively, the type-(a) piece of Λ^5, and the pieces of $-\Sigma(p)\gamma_5$ and $-\gamma_5\Sigma(p')$ corresponding to the type-(a) piece of Λ_μ^5 in Eq. (8). Summing over all type-(a) contributions to Λ_μ^5, we get

$$(p-p')^\mu\Lambda_\mu^{5(a)}(p,p') = 2m_0\Lambda^{5(a)}(p,p') - \Sigma(p)\gamma_5 - \gamma_5\Sigma(p')\,. \tag{11}$$

We turn next to contributions to Λ_μ^5 of type (b). A

typical term is

$$\int d^4r\,\mathrm{Tr}\left\{\sum_{k=1}^{2n}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\gamma^{(k)}\frac{1}{r+p_k-m_0}\gamma_\mu\gamma_5\right.$$
$$\left.\times\frac{1}{r+p_k+p'-p-m_0}\prod_{j=k+1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j+p'-p-m_0}\right]\right\}\times(\cdots). \quad (12)$$

Multiplying by $(p-p')^\mu$ and using Eq. (9) gives

$$\int d^4r\,\mathrm{Tr}\left\{\sum_{k=1}^{2n}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\gamma^{(k)}\frac{1}{r+p_k-m_0}2m_0\gamma_5\right.$$
$$\left.\times\frac{1}{r+p_k+p'-p-m_0}\prod_{j=k+1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j+p'-p-m_0}\right]\right\}$$
$$\times(\cdots)+\int d^4r\,\mathrm{tr}\left\{\gamma_5\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\right.$$
$$\left.-\gamma_5\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j+p'-p-m_0}\right]\right\}(\cdots). \quad (13)$$

The first term in Eq. (13) is the type-(b) contribution to Λ^5 corresponding to Eq. (12), while making the change of variable $r\to r+p'-p$ in the integration in the second term causes the second and third terms to cancel. This gives, when we sum over all type-(b) contributions,

$$(p-p')^\mu\Lambda_\mu{}^{5(b)}(p,p')=2m_0\Lambda^{5(b)}(p,p'). \quad (14)$$

The Ward identity of Eq. (7) is finally obtained by adding Eqs. (11) and (14).

Clearly, the only step of the above derivation which is not simply an algebraic rearrangement is the *change of integration variable* in the second term of Eq. (13). This will be a valid operation provided that the integral is at worst superficially logarithmically divergent, a condition that is satisfied by loops with four or more photons, that is, loops with $n\geq 2$. However, when the loop is a triangle graph with only two photons emerging (See Fig. 2) we have $n=1$, and the integral in Eq. (13) appears to be quadratically divergent. Actually, since

$$\mathrm{tr}\{\gamma_5\gamma^{(1)}r\gamma^{(2)}r\}=0, \quad (15)$$

the integral in the $n=1$ case is superficially *linearly* divergent. Since it is well known that translation of a linearly divergent integral is not necessarily a valid operation,[2] we must check carefully to see whether Eq. (14) holds for the triangle graph.

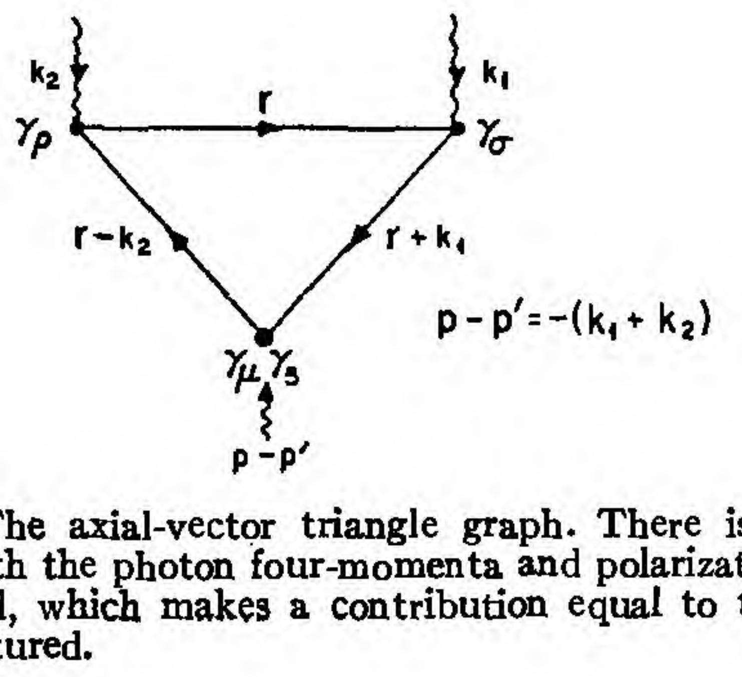


FIG. 2. The axial-vector triangle graph. There is a second diagram, with the photon four-momenta and polarization indices interchanged, which makes a contribution equal to that of the diagram pictured.

To do this we make use of an explicit expression for the triangle graph calculated by Rosenberg.[3] The sum of the diagram illustrated in Fig. 2 and the corresponding diagram with the two photons interchanged is

$$\frac{-ie_0^2}{(2\pi)^4}R_{\sigma\rho\mu}\equiv 2\int\frac{d^4r}{(2\pi)^4}(-1)\mathrm{tr}\left\{\frac{i}{r+k_1-m_0}(-ie_0\gamma_\sigma)\right.$$
$$\left.\times\frac{i}{r-m_0}(-ie_0\gamma_\rho)\frac{i}{r-k_2-m_0}\gamma_\mu\gamma_5\right\}. \quad (16)$$

Evaluation of Eq. (16) by the usual regulator techniques leads to the following expression for $R_{\sigma\rho\mu}$ [A_j denotes $A_j(k_1,k_2)$]:

$$\begin{aligned}R_{\sigma\rho\mu}(k_1,k_2)={}&A_1k_1{}^\tau\epsilon_{\tau\sigma\rho\mu}+A_2k_2{}^\tau\epsilon_{\tau\sigma\rho\mu}\\&+A_3k_{1\rho}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\sigma\mu}+A_4k_{2\rho}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\sigma\mu}\\&+A_5k_{1\sigma}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\rho\mu}+A_6k_{2\sigma}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\rho\mu},\\A_1={}&k_1\cdot k_2A_3+k_2{}^2A_4,\\A_2={}&k_1{}^2A_5+k_1\cdot k_2A_6,\\A_3(k_1,k_2)={}&-A_6(k_2,k_1)=-16\pi^2I_{11}(k_1,k_2),\\A_4(k_1,k_2)={}&-A_5(k_2,k_1)=16\pi^2[I_{20}(k_1,k_2)-I_{10}(k_1,k_2)],\end{aligned} \quad (17)$$

where

$$I_{st}(k_1,k_2)=\int_0^1dx\int_0^{1-x}dy\,x^sy^t[y(1-y)k_1{}^2+x(1-x)k_2{}^2+2xyk_1\cdot k_2-m_0{}^2]^{-1}. \quad (18)$$

---

[2] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Co., Inc., Cambridge, Mass., 1955), pp. 458–461.

[3] L. Rosenberg, Phys. Rev. 129, 2786 (1963). In Eq. (16) and Fig. 2, we have labeled the legs of the triangle in accordance with Rosenberg's notation, which differs from the labeling convention used in Eqs. (12) and (13). Because the integral defining the triangle graph is linearly divergent, the value of the triangle graph is ambiguous and depends on the labeling convention and the method of evaluation of the integral. For example, if Eq. (16) is evaluated by symmetric integration about the origin in r space, the value of $R_{\sigma\rho\mu}$ so obtained satisfies the usual axial-vector Ward identity (but is not gauge-invariant with respect to the vector indices). If, on the other hand, Eq. (16) is evaluated by symmetric integration around some other point in r space, say $r=k_1$ [or, alternatively, if we integrate symmetrically around $r=0$ but label the triangle using the convention of Eqs. (12) and (13)], then the result has an anomalous axial-vector Ward identity. The value in Eq. (17) which we have assigned to $R_{\sigma\rho\mu}$ is the unique value which is gauge-invariant with respect to the vector indices. Further discussion of the ambiguity in the definition of Eq. (16), and a justification of the specific choice in Eq. (17), are given in the Appendix.

We will also need an expression for the triangle graph with $\gamma_\mu\gamma_5$ replaced by $2m_0\gamma_5$. Defining

$$\frac{-ie_0^2}{(2\pi)^4}2m_0R_{\sigma\rho}\equiv 2\int\frac{d^4r}{(2\pi)^4}(-1)\mathrm{tr}\left\{\frac{i}{r+k_1-m_0}(-ie_0\gamma_\sigma)\times\frac{i}{r-m_0}(-ie_0\gamma_\rho)\frac{i}{r-k_2-m_0}2m_0\gamma_5\right\},\qquad(19)$$

we find that

$$R_{\sigma\rho}=k_1^{\xi}k_2^{\tau}\epsilon_{\xi\tau\sigma\rho}B_1\,,\qquad B_1=8\pi^2m_0I_{00}(k_1,k_2)\,.\qquad(20)$$

We are now ready to calculate the divergence of the axial-vector triangle diagram. If the Ward identity holds, we should find

$$-(k_1+k_2)^\mu R_{\sigma\rho\mu}=2m_0R_{\sigma\rho}\,,\qquad(21)$$

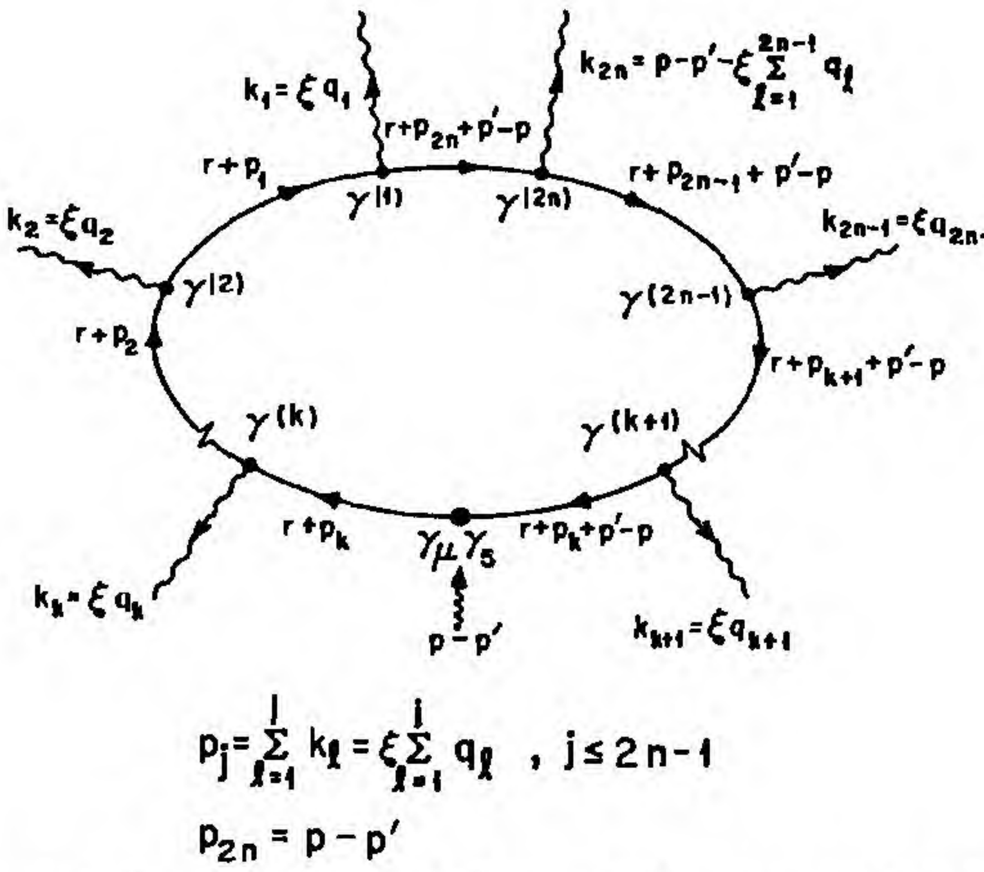


FIG. 3. Diagram for calculation of the asymptotic behavior of the general axial-vector loop.

but from Eqs. (16)–(20) we find, instead,

$$-(k_1+k_2)^\mu R_{\sigma\rho\mu}=2m_0R_{\sigma\rho}+8\pi^2k_1^{\xi}k_2^{\tau}\epsilon_{\xi\tau\sigma\rho}\,.\qquad(22)$$

We see that *the axial-vector Ward identity fails in the case of the triangle graph.* The failure is a result of the fact that the integration variable in a linearly divergent Feynman integral cannot be freely translated.

The breakdown of the axial-vector Ward identity which we have just found is related to another anomalous property of the triangle graph. To see this, let us consider the behavior of the general axial-vector loop diagram with $2n$ photon vertices (See Fig. 3), as the $2n-1$ independent photon momenta $k_1,\cdots,k_{2n-1}$ approach infinity simultaneously in the manner

$$k_j=\xi q_j\,,\quad j=1,\cdots,2n-1\,;\qquad q_j\text{ fixed}\,,\quad \xi\to\infty\,,\qquad(23)$$

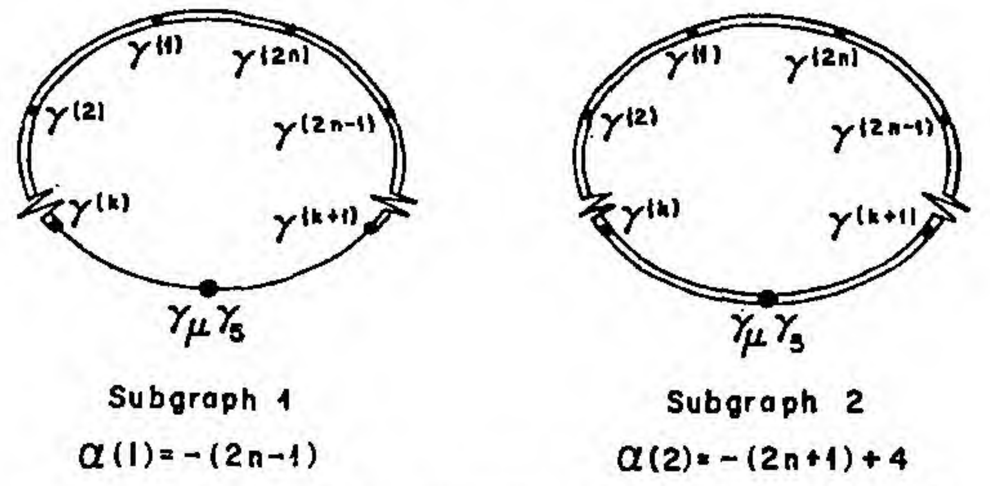


FIG. 4. Subgraphs (doubled lines) which determine the asymptotic behavior of Fig. 3.

while the momentum $p-p'$ carried by the axial-vector current is held fixed. According to Weinberg's theorem,[4] the asymptotic behavior of the loop graph in this limit is

$$\xi^\alpha(\ln\xi)^\beta\,,\qquad(24)$$

where β is undetermined by Weinberg's analysis and where α is the maximum of the superficial divergences[5] $\alpha(g)$ of the subgraphs[5] g linking the $2n$ photon lines (i.e., linking the momenta which are becoming infinite). For the diagram of Fig. 3 there are two such subgraphs, illustrated in Fig. 4, with superficial divergences $\alpha(1)=-2n+1$ and $\alpha(2)=-2n+3$. Thus, the asymptotic coefficient α is $\alpha(2)=-2n+3$, and comes from the subgraph in which all propagators in the loop are involved. Now Weinberg's theorem always tells us what the maximal asymptotic *power* of a graph is, but it does not guarantee that the *coefficient* of the maximal term is nonvanishing. In fact, in the case of the axial-vector loop diagram we will show that the coefficient of the $\xi^{-2n+3}(\ln\xi)^\beta$ term *does* vanish, so that the leading asymptotic behavior is $\xi^{-2n+2}(\ln\xi)^{\beta'}$, one power lower than is predicted by naive power counting. Let us denote by $L(p-p',m_0;p_1,\cdots,p_{2n-1})$ the graph illustrated in Fig. 3,

$$L(p-p',m_0;p_1,\cdots,p_{2n-1})=\int d^4r\,\mathrm{Tr}\left\{\sum_{k=1}^{2n}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\times\gamma^{(k)}\frac{1}{r+p_k-m_0}\gamma_\mu\gamma_5\frac{1}{r+p_k+p'-p-m_0}\times\prod_{j=k+1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j+p'-p-m_0}\right]\right\}.\qquad(25)$$

[4] S. Weinberg, Phys. Rev. 118, 838 (1960). For a simplified exposition of Weinberg's results, see J. D. Bjorken and S. D. Drell, Ref. 1, pp. 317–330 and pp. 364–368. Weinberg's theorem applies for arbitrary *spacelike* four-vectors q_j. There can also be powers of ln lnξ, ln ln lnξ, etc., in Eq. (24), which we do not indicate explicitly.

[5] The superficial divergence of the subgraph is obtained, as usual, by adding -1 for each internal fermion line, -2 for each internal boson line, and $+4$ for each internal integration. For the precise definition of subgraph in the general case, see Ref. 4.

Clearly we can write

$$L(p-p',m_0;\, p_1,\cdots,p_{2n-1})$$
$$\text{(A)}\quad =L(p-p',m_0;\, p_1,\cdots,p_{2n-1}) -L(0,m_0;\, p_1,\cdots,p_{2n-1}) \qquad (26)$$
$$\text{(B)}\quad +L(0,m_0;\, p_1,\cdots,p_{2n-1})-L(0,0;\, p_1,\cdots,p_{2n-1})$$
$$\text{(C)}\quad +L(0,0;\, p_1,\cdots,p_{2n-1}).$$

Because differencing the loop graph with respect to either the axial-vector current four-momentum $p-p'$ or the fermion mass m_0 decreases the degree of divergence by one, terms (A) and (B) on the right-hand side of Eq. (26) have $\alpha(2)=-2n+2$, and therefore behave asymptotically as $\xi^{-2n+2}(\ln\xi)^{\beta'}$. Term (C) on the right-hand side of Eq. (26) can be rewritten as

$$L(0,\,0;\, p_1,\,\cdots,\, p_{2n-1})$$
$$=\int d^4r\,\mathrm{Tr}\left\{\sum_{k=1}^{2n}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{r+p_j}\right]\gamma^{(k)}\frac{1}{r+p_k}\gamma_\mu\gamma_5 \times\frac{1}{r+p_k}\prod_{j=k+1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j}\right]\right\}$$
$$=\int d^4r\,\mathrm{Tr}\left\{\gamma_5\frac{\partial}{\partial r^\mu}\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j}\right]\right\}. \qquad (27)$$

Integrating by parts with respect to r gives

$$L(0,0;\, p_1,\cdots,p_{2n-1})=0,$$

proving that the asymptotic behavior of the loop graph is one power better than given by Weinberg's theorem.

The only nonalgebraic step in this proof is the integration by parts with respect to r, an operation which is valid provided that the integration variable in

$$\int d^4r\,\mathrm{Tr}\left\{\gamma_5\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j}\right]\right\} \qquad (28)$$

can be freely translated. This is the same condition as we found above for validity of the axial-vector Ward identity. Thus again, our proof is valid for $n\geq 2$, but we expect possible trouble in the case of the triangle graph ($n=1$). From the explicit expression for the triangle graph in Eqs. (17) and (18), we see that if we write $k_1=\xi q$, $k_2=-\xi q+p'-p$, then as $\xi\to\infty$ we find

$$R_{\sigma\rho\mu}(k_1,k_2)\to -8\pi^2\xi q^\tau\epsilon_{\tau\sigma\rho\mu}+O(\ln\xi). \qquad (29)$$

In other words, the asymptotic power is $\alpha=1=-2n+3$, as given by Weinberg's rules, rather than one power lower, as is the case for the loop graphs with $n\geq 2$. It is easy to check that when Eq. (29) is multiplied by $-(k_1+k_2)^\mu$, the term with the anomalous asymptotic behavior agrees, for large ξ, with the term in Eq. (22) which violates the Ward identity. Thus, the breakdown of the axial-vector Ward identity in the triangle graph

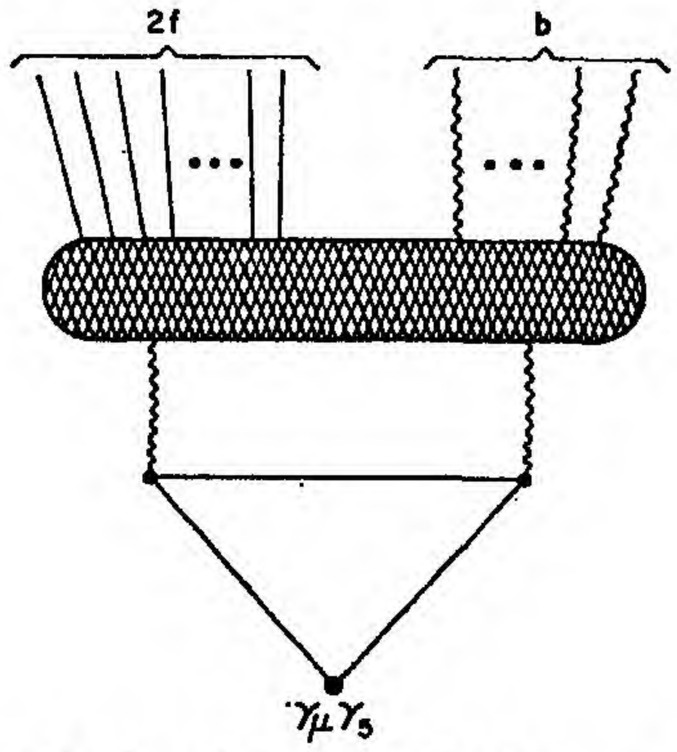


FIG. 5. Contribution of the triangle diagram to the general axial-vector vertex. We have not drawn the second diagram in which the photon lines emerging from the triangle are crossed.

and the anomalous asymptotic behavior of the triangle graph are basically the same phenomenon.

It is clear that the breakdown of the Ward identity for the basic triangle graph will also cause failure of the Ward identity for any graph of the type illustrated in Fig. 5, in which the two photon lines coming out of the triangle graph join onto a "blob" from which $2f$ fermion and b boson lines emerge. From Eq. (22) for the divergence of the basic triangle graph, it is possible to show that the breakdown of the axial-vector Ward identity in the general case is simply described by replacing Eq. (4) for the axial-vector-current divergence (which we have shown to be incorrect) by

$$\frac{\partial}{\partial x_\mu}j_\mu{}^5(x)=2im_0 j^5(x)+\frac{\alpha_0}{4\pi}:F^{\xi\sigma}(x)F^{\tau\rho}(x):\,\epsilon_{\xi\sigma\tau\rho}. \qquad (30)$$

[Equation (30) is easily verified by using the Feynman rules for the vertices of $j_\mu{}^5$, j^5, and $(\alpha_0/4\pi):F^{\xi\sigma}F^{\tau\rho}:\,\epsilon_{\xi\sigma\tau\rho}$, which are given in Fig. 6.] For example, if we define $\bar{F}(p,p')$ by

$$S_F'(p)\bar{F}(p,p')S_F'(p')=-\int d^4x d^4y\, e^{ip\cdot x}e^{-ip'\cdot y} \times\langle T(\psi(x):F^{\xi\sigma}(0)F^{\tau\rho}(0):\epsilon_{\xi\sigma\tau\rho}\bar{\psi}(y))\rangle_0, \qquad (31)$$

then the axial-vertex Ward identity of Eq. (5) is modified to read

$$(p-p')^\mu\Gamma_\mu{}^5(p,p')=2m_0\Gamma^5(p,p')-i(\alpha_0/4\pi)\bar{F}(p,p') +S_F'(p)^{-1}\gamma_5+\gamma_5 S_F'(p')^{-1}. \qquad (32)$$

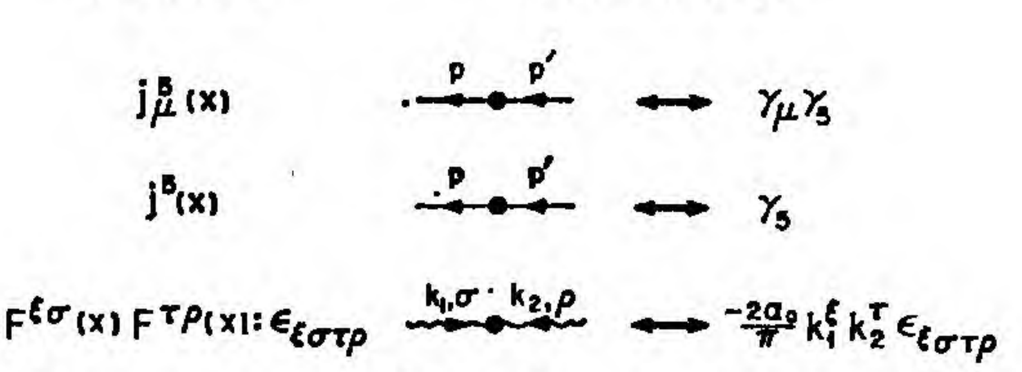


FIG. 6. Feynman rules for the vertices appearing in Eq. (30).

Equation (30), which is the principal result of this section, states the surprising fact that *the axial-vector-current divergence, as calculated in perturbation theory, contains a well-defined extra term which is not obtained when the axial-vector divergence is calculated by formal use of the equations of motion.*[6]

II. CONSEQUENCES OF THE EXTRA TERM

In this section we investigate the consequences of the extra term which we have found in the axial-vector-current divergence [Eq. (30)] and in the axial-vector-current Ward identity [Eq. (32)]. We consider, in particular, the questions of (A) renormalization of the axial-vector vertex, (B) radiative corrections to $\nu_l l$ scattering, and (C) the connection between γ_5 invariance and a conserved axial-vector current in massless quantum electrodynamics.

A. Renormalization of the Axial-Vector Vertex

Recently, Preparata and Weisberger[7] have proved the following theorem: If a local current, constructed as a bilinear product of fermion fields, is conserved apart from mass terms, then the vertex parts of both the current and its divergence are made finite by multiplication by the wave-function renormalization constants of the fields from which the current is constructed. *If* Eq. (4) correctly described the divergence of the axial-vector current in spinor electrodynamics, then the theorem of Preparata and Weisberger would apply in this case. However, we have seen that the divergence is actually given by Eq. (30), and involves an additional term which is *not* a mass term. The effect of this extra term, we shall see, is to cause the Preparata-Weisberger argument to break down.

First let us review how the Preparata-Weisberger result could be derived if Eq. (4), and the corresponding Ward identify of Eq. (5), were true. Since both $j_\mu{}^5$ and j^5 are local bilinear products of fermion fields, the vertex parts $\Gamma_\mu{}^5$ and Γ^5 are *multiplicatively renormalizable.* Thus we can write

$$\begin{aligned}\Gamma_\mu{}^5(p,p')&=Z_A{}^{-1}\tilde{\Gamma}_\mu{}^5(p,p'),\\ \Gamma^5(p,p')&=Z_D{}^{-1}\tilde{\Gamma}^5(p,p'),\\ S_F'(p)&=Z_2\tilde{S}_F'(p),\end{aligned}\tag{33}$$

where the tilde quantities are finite (cutoff-independent) and where Z_A, Z_D, and Z_2 are cutoff-dependent renormalization constants. Substituting Eq. (32) into Eq. (5) we get

$$(p-p')^\mu\tilde{\Gamma}_\mu{}^5(p,p')=(2m_0Z_A/Z_D)\tilde{\Gamma}^5(p,p')+(Z_A/Z_2)[\tilde{S}_F'(p)^{-1}\gamma_5+\gamma_5\tilde{S}_F'(p')^{-1}],\tag{34}$$

[6] We show in the Appendix that this extra term cannot be eliminated by redefining the triangle graph.

[7] G. Preparata and W. I. Weisberger, Phys. Rev. 175, 1965 (1968), Appendix C.

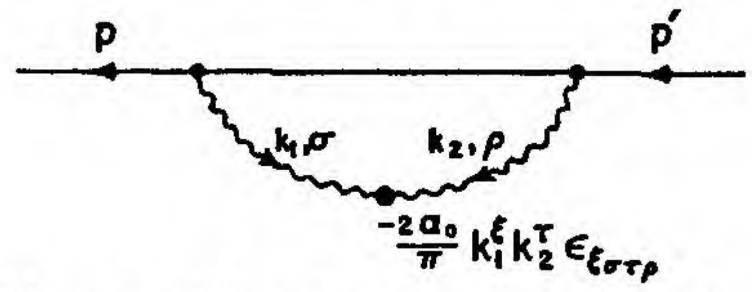


FIG. 7. Diagram giving the lowest-order contribution of the extra term in Eq. (32). The heavy dot denotes the vertex of $(\alpha_0/4\pi):F^{\xi\sigma}F^{\tau\rho}:\epsilon_{\xi\sigma\tau\rho}$.

and varying the cutoff gives

$$0=\delta(2m_0Z_A/Z_D)\tilde{\Gamma}^5(p,p')+\delta(Z_A/Z_2)\times[\tilde{S}_F'(p)^{-1}\gamma_5+\gamma_5\tilde{S}_F'(p')^{-1}].\tag{35}$$

Putting p,p', or both on mass shell then implies that

$$\delta(2m_0Z_A/Z_D)=\delta(Z_A/Z_2)=0,\tag{36}$$

which means that both $2m_0Z_A/Z_D$ and Z_A/Z_2 are cutoff-independent, and hence finite. Thus, if Eqs. (4) and (5) were correct, multiplication by the wave-function renormalization constant Z_2 would make $\Gamma_\mu{}^5$ and Γ^5 fiinte.

Let us now consider the actual situation, in which the divergence of the axial-vector current is given by Eq. (30) and the axial-vector Ward identity by Eq. (32). The extra term in Eq. (32) first appears in order $\alpha_0{}^2$ of perturbation theory. [See Fig. 7.] This lowest-order contribution is already logarithmically divergent; introducing a cutoff by replacing the photon propagator $1/(q^2+i\epsilon)$ with $[1/(q^2+i\epsilon)][-\Lambda^2/(-\Lambda+q^2+i\epsilon)]$, we find that

$$-i(\alpha_0/4\pi)\bar{F}(p,p')=-\tfrac{3}{4}(\alpha_0/\pi)^2\ln(\Lambda^2/m^2)(p-p)^\mu\times\gamma_\mu\gamma_5+\alpha_0{}^2\times\text{finite}+O(\alpha_0{}^3).\tag{37}$$

We will also need part of the expression for $\Gamma^5(p,p')$ to order α_0,

$$\begin{aligned}\Gamma^5(p,p')&=\gamma_5[1+O(\alpha_0)]+(\alpha_0/2\pi)m_0\times I(p,p')(p-p')^\mu\gamma_\mu\gamma_5+O(\alpha_0{}^2),\\ I(p,p')&=\int_0^1dx\int_0^{1-x}dy[x(1-x)p^2+y(1-y)p'^2-2xyp\cdot p'-(x+y)m_0{}^2]^{-1}.\end{aligned}\tag{38}$$

Comparing Eqs. (37) and (38), we see that *it is impossible to cancel away the divergence in Eq. (37) by adding to it a constant multiple of Eq. (38)*: A constant counter term of order $\alpha_0{}^2$ multiplying the leading γ_5 term in Eq. (38) cannot cancel the divergence in Eq. (37), because the latter is proportional to $(p-p')^\mu\gamma_\mu\gamma_5$, while a constant counter term of order α_0 multiplying the $(p-p')^\mu\gamma_\mu\gamma_5$ term in Eq. (38) cannot cancel the divergence in Eq. (37) because of the nontrivial functional dependence of $I(p,p')$ on p and p'. In other words, the axial-vector divergence with the extra term included,

$$2m_0\Gamma^5(p,p')-i(\alpha_0/4\pi)\bar{F}(p,p'),\tag{39}$$

is *not multiplicatively renormalizable.*

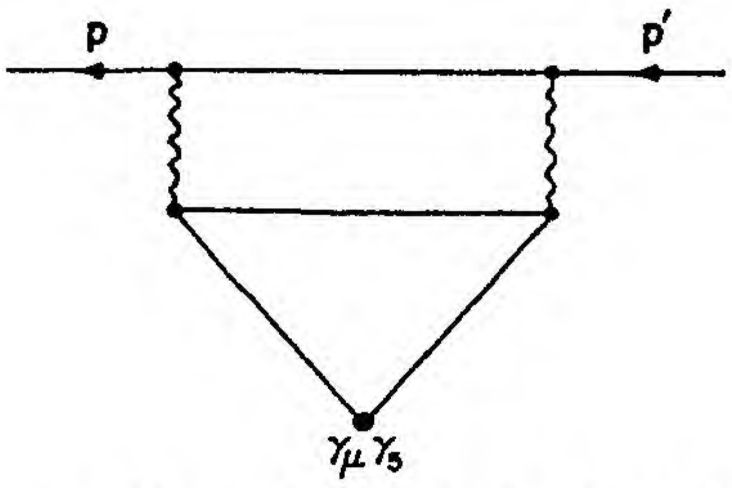


FIG. 8. Lowest-order contribution of the triangle diagram to the axial-vector vertex. We have not drawn the diagram in which the photon lines are crossed.

Since multiplicative renormalizability of the divergence was essential to the Preparata-Weisberger argument outlined above, this argument no longer applies. We expect, then, that even after multiplication by Z_2, there will still be logarithmically divergent terms in the axial-vector vertex. Such terms first appear in order α_0^2 of perturbation theory, as a result of the diagram shown in Fig. 8; the divergence of Fig. 8 is just a consequence of the anomalous asymptotic behavior of the triangle graph pointed out in Sec. I. Introducing a cutoff in the photon propagator as above, we find that

$$Z_2\Gamma_\mu^5(p,p')=\gamma_\mu\gamma_5[1-\tfrac{3}{4}(\alpha_0/\pi)^2\ln(\Lambda^2/m^2)] +\alpha_0\times\text{finite}+\alpha_0^2\times\text{finite}+O(\alpha_0^3). \quad (40)$$

Equation (40) shows explicitly that the axial-vector vertex, while still multiplicatively renormalizable, is not simply made finite by multiplication by the wave-function renormalization constant Z_2. Rather, we have [see Eq. (33)]

$$Z_A=Z_2[1+\tfrac{3}{4}(\alpha_0/\pi)^2\ln(\Lambda^2/m^2)+O(\alpha_0^3)]. \quad (41)$$

B. Radiative Corrections to $\nu_l l$ Scattering

As an application of Eq. (40), let us consider the radiative corrections to $\nu_l l$ scattering, where l is a μ or an e. According to the usual local current-current theory, the leptonic weak interactions are described by the effective Lagrangian

$$\mathcal{L}_{\text{eff}}=(G/\sqrt{2})j_\lambda^\dagger j^\lambda, \quad (42)$$

where $G\approx10^{-5}/M_{\text{proton}}^2$ is the Fermi constant and where[8]

$$j^\lambda=\bar{\nu}_\mu\gamma^\lambda(1-\gamma_5)\mu+\bar{\nu}_e\gamma^\lambda(1-\gamma_5)e \quad (43)$$

is the leptonic current. In addition to the usual terms describing muon decay, Eq. (42) contains the terms

$$(G/\sqrt{2})[\bar{\mu}\gamma_\lambda(1-\gamma_5)\nu_\mu\bar{\nu}_\mu\gamma^\lambda(1-\gamma_5)\mu +\bar{e}\gamma_\lambda(1-\gamma_5)\nu_e\bar{\nu}_e\gamma^\lambda(1-\gamma_5)e], \quad (44)$$

which describe elastic neutrino-lepton scattering. In order to study radiative corrections to the basic $\nu_l l$ scattering process, it is convenient to use a Fierz transformation to rewrite Eq. (44) in the form (the so-called "charge retention ordering")

$$(G/\sqrt{2})[\bar{\mu}\gamma_\lambda(1-\gamma_5)\mu\bar{\nu}_\mu\gamma^\lambda(1-\gamma_5)\nu_\mu +\bar{e}\gamma_\lambda(1-\gamma_5)e\bar{\nu}_e\gamma^\lambda(1-\gamma_5)\nu_e]. \quad (45)$$

The radiative corrections to Eq. (45) may then be obtained simply by calculating the radiative corrections to the charged lepton currents $\bar{\mu}\gamma_\lambda(1-\gamma_5)\mu$ and $\bar{e}\gamma_\lambda(1-\gamma_5)e$, without any reference to the neutrino currents.

Now, application of standard electrodynamic perturbation theory shows that the effect of the radiative corrections to the charged lepton currents is to replace the matrix elements $\bar{\mu}\gamma_\lambda(1-\gamma_5)\mu$, $\bar{e}\gamma_\lambda(1-\gamma_5)e$ (we use μ, e to denote spinors here) by

$$\bar{\mu}Z_2^{(\mu)}[\Gamma_\lambda^{(\mu)}-\Gamma_\lambda^{5(\mu)}]\mu,\quad \bar{e}Z_2^{(e)}[\Gamma_\lambda^{(e)}-\Gamma_\lambda^{5(e)}]e. \quad (46)$$

In Eq. (46), $\Gamma_\lambda^{(\mu,e)}$ and $\Gamma_\lambda^{5(\mu,e)}$ denote the proper vector and axial-vector vertices, while the wave-function renormalization factors $Z_2^{(\mu,e)}$ come from self-energy insertions on the external lepton lines which run into and out of the proper vertices. From the usual electrodynamic Ward identity for the vector part, we know that $Z_2^{(\mu)}\Gamma_\lambda^{(\mu)}$ and $Z_2^{(e)}\Gamma_\lambda^{(e)}$ are finite. On the other hand, Eq. (40) tells us that

$$Z_2^{(\mu,e)}\Gamma_\lambda^{5(\mu,e)}=\gamma_\lambda\gamma_5[1-\tfrac{3}{4}(\alpha_0/\pi)^2\ln(\Lambda^2/m^2)] +\alpha_0\times\text{finite}+\alpha_0^2\times\text{finite}+O(\alpha_0^3), \quad (47)$$

which means that, on account of the presence of axial-vector triangle diagrams, *the radiative corrections to $\nu_e e$ and $\nu_\mu\mu$ scattering diverge in the fourth order of perturbation theory.* This result contrasts sharply with the fact that the radiative corrections to muon decay or to the scattering reaction $\nu_\mu+e\to\nu_e+\mu$ are finite to all orders in perturbation theory.[7] The crucial difference between the two cases, of course, is that because of separate muon and electron-number conservation, the current $\bar{\mu}\gamma_\lambda(1-\gamma_5)e$ cannot couple into closed electron or muon loops, and thus the troublesome triangle diagram is not present.

Two points of view can be taken towards the divergent radiative corrections in $\nu_l l$ scattering. One viewpoint is that we know, in any case, that the local current-current theory of leptonic weak interactions cannot be correct, since this theory leads at high energies to nonunitary matrix elements, and since it gives divergent results for higher-order weak-interaction effects.[9] Thus, it is entirely possible that the modifications in Eq. (44) necessary to give a satisfactory weak-interaction theory will also cure the disease of infinite radiative corrections in $\nu_l l$ scattering. The other viewpoint is that we should try to make the radiative corrections to $\nu_l l$ scattering finite, within the framework of a local weak-interaction theory. It turns out that this

[8] We omit the normal ordering signs.

[9] For recent discussions of the sicknesses of the local current-current theory and their possible remedies, see N. Christ, Phys. Rev. 176, 2086 (1968); and M. Gell-Mann, M. L. Goldberger, N. M. Kroll, and F. E. Low, Phys. Rev. (to be published).

is possible, if we introduce $\nu_e\mu$ and $\nu_\mu e$ scattering terms into the effective Lagrangian, so that Eq. (44) is replaced by

$$(G/\sqrt{2})[\bar{\mu}\gamma_\lambda(1-\gamma_5)\mu-\bar{e}\gamma_\lambda(1-\gamma_5)e] \times[\bar{\nu}_\mu\gamma^\lambda(1-\gamma_5)\nu_\mu-\bar{\nu}_e\gamma^\lambda(1-\gamma_5)\nu_e]. \quad (48)$$

This works because the troublesome extra term in Eq. (30) is independent of the bare mass m_0, so that it cancels between the muon and electron terms in Eq. (48), giving[10]

$$\frac{\partial}{\partial x_\lambda}[\bar{\mu}\gamma_\lambda\gamma_5\mu-\bar{e}\gamma_\lambda\gamma_5 e]=2im_0^{(\mu)}\bar{\mu}\gamma_5\mu-2im_0^{(e)}\bar{e}\gamma_5 e. \quad (49)$$

Application of the Preparata-Weisberger argument to Eq. (49) then shows that the radiative corrections to Eq. (48) are finite in all orders of perturbation theory. Experimentally, it will be possible to distinguish between Eq. (48) and Eq. (44) by looking for elastic scattering of muon neutrinos from electrons.

C. Connection Between γ_5 Invariance and a Conserved Axial-Vector Current in Massless Electrodynamics

Finally, let us discuss the effects of the axial-vector triangle diagram in the case of massless spinor electrodynamics [Eq. (1) with $m_0=0$]. We will find that the triangle diagram leads to a breakdown of the usual connection between symmetries of the Lagrangian and conserved currents. As in our previous discussions, we begin by describing the standard theory, which holds in the absence of singular phenomena.[8] Let $\{\Phi(x)\}=\{\Phi_1(x),\Phi_2(x),\cdots\}$ and $\{\partial_\lambda\Phi\}$ be a set of canonical fields and their space-time derivatives, and let us consider the field theory described by the Lagrangian density

$$\mathcal{L}(x)\equiv\mathcal{L}[\{\Phi\},\{\partial_\lambda\Phi\}]. \quad (50)$$

To establish the connection between invariance properties of $\mathcal{L}$ and conserved currents, we make the infinitesimal, local gauge transformation on the fields,

$$\Phi_j(x)\rightarrow\Phi_j(x)+\Lambda(x)G_j[\{\Phi(x)\}], \quad (51)$$

and define the associated current J^α by

$$J^\alpha=-\delta\mathcal{L}/\delta(\partial_\alpha\Lambda). \quad (52)$$

Then, by using the Euler-Lagrange equations of motion of the fields, we easily find[11] that the divergence of the current is given by

$$\partial_\alpha J^\alpha=-\delta\mathcal{L}/\delta\Lambda. \quad (53)$$

[10] What is happening here is that the muon triangle diagram and the electron triangle diagram contribute with opposite sign, and so regularize each other.

[11] For details, see S. L. Adler and R. F. Dashen, *Current Algebras* (W. A. Benjamin, Inc., New York, 1968), pp. 15–18.

In particular, if the gauge transformation of Eq. (51), with *constant* gauge function Λ, leaves the Lagrangian invariant, then $\delta\mathcal{L}/\delta\Lambda=0$ and the current J^α is conserved. Thus, to any continuous invariance transformation of the Lagrangian there is associated a conserved current. It is also easily verified that the charge $Q(t)=\int d^3x\,J^0(\mathbf{x},t)$ associated with the current J^α has the properties

$$dQ(t)/dt=0, \quad (54a)$$

$$[Q,\Phi_j(x)]=iG_j(x). \quad (54b)$$

Equation (54b) states that Q is the generator of the gauge transformation in Eq. (51), for constant Λ.

Let us now specialize to the case of massless electrodynamics, with Eq. (51) the gauge transformation

$$\psi(x)\rightarrow[1+i\gamma_5\Lambda(x)]\psi(x). \quad (55)$$

When Λ is a constant and $m_0=0$, this transformation leaves the Lagrangian of Eq. (1) invariant, so that according to Eq. (53), the associated current J^α should be conserved. But calculating J^α, we find

$$J^\alpha=-\delta\mathcal{L}/\delta(\partial_\alpha\Lambda)=\bar{\psi}\gamma^\alpha\gamma_5\psi, \quad (56)$$

which according to Eq. (30) has the divergence

$$\partial_\alpha J^\alpha=(\alpha_0/4\pi)F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho}. \quad (57)$$

Thus, Eq. (53), which was obtained by formal calculation using the equations of motion, breaks down in this case. We see that because of the presence of the axial-vector triangle diagram, *even though the Lagrangian (and all orders of perturbation theory) of massless electrodynamics are γ_5 invariant, the axial-vector current associated with the γ_5 transformation is not conserved.*

However, it is amusing that even though there is no conserved current connected with the γ_5 transformation, there is still a generator $\tilde{Q}^5$ with the properties of Eq. (54). To see this, let us consider the quantity $\tilde{j}^5$ defined by

$$\tilde{j}_\mu{}^5(x)=j_\mu{}^5(x)-\frac{\alpha_0}{\pi}A^\xi(x)\frac{\partial A^\tau(x)}{\partial x_\rho}\epsilon_{\xi\mu\tau\rho}; \quad (58)$$

referring to Eq. (30), we see that

$$\frac{\partial}{\partial x_\mu}\tilde{j}_\mu{}^5(x)=0. \quad (59)$$

Although $\tilde{j}_\mu{}^5$ is conserved, it is explicitly *gauge-dependent* and therefore *is not an observable current operator.* But the associated charge

$$\tilde{Q}^5=\int d^3x\,\tilde{j}_0{}^5(x) =\int d^3x\left[\psi^\dagger(x)\gamma_5\psi(x)+\frac{\alpha_0}{\pi}\mathbf{A}\cdot\nabla\times\mathbf{A}\right] \quad (60)$$

is gauge-invariant and therefore observable. According to Eq. (59), $\bar{Q}^5$ is time-independent, and its commutator with $\psi(x)$ (calculated formally by use of the canonical commutation relations) is

$$[\bar{Q}^5,\psi(x)]=-\gamma_5\psi(x)=i[i\gamma_5\psi(x)]. \tag{61}$$

Comparison with Eq. (59) then shows that $\bar{Q}^5$ is the conserved generator of the γ_5 transformations.[12]

After this manuscript was completed, we learned that Bell and Jackiw[13] had independently studied the anomalous properties of the axial-vector triangle graph, in the context of the σ model. In the Appendix we discuss certain questions raised both by the paper of Bell and Jackiw and in conversations with Professor S. Coleman.

Note added in proof. (1) All field quantities appearing in the paper denote *unrenormalized* fields, with the one exception that in Eqs. (A29), (A30), and (A34), ϕ_{π^0} and ϕ_η denote, respectively, the renormalized pion and η fields.

(2) It is our claim that Eq. (30) is an *exact* result, valid to all orders in electromagnetism, and similarly that the σ-model analog, Eq. (A22), is exact to all orders in both the electromagnetic and strong couplings. These conclusions follow in our diagrammatic analysis from the fact that electromagnetic or strong radiative corrections to the basic triangle always involve axial-vector loops with more than three vertices, which satisfy the normal axial-vector Ward identities. A more detailed discussion of this question will be given by the author and W. A. Bardeen (to be published).

(3) Field-theoretic derivations of Eq. (30) have been given by C. R. Hagen (to be published), R. Jackiw and K. Johnson (to be published), B. Zumino (to be published), and R. A. Brandt (to be published). Jackiw and Johnson point out that the essential features of the field-theoretic derivation, in the case of external electromagnetic fields, are contained in J. Schwinger, Phys. Rev. **82**, 664 (1951).

(4) In Eq. (A1) we state that the general form of the triangle diagram is $R_{\sigma\rho\mu}$, Rosenberg's gauge-invariant expression, plus an arbitrary multiple of $\epsilon_{\tau\sigma\rho\mu}(k_1-k_2)^\tau$; we infer this form for the extra term by studying how the triangle graph is changed by shifts in the integration variable. It is easy to see that this is the *only allowed form* for the ambiguity, by noting that the extra term must satisfy the following conditions. (i) The extra term must have the dimensions of a mass; (ii) the extra term must be a three-index ($\sigma\rho\mu$) Lorentz pseudotensor; (iii) the extra term must be symmetric under interchange of the photon variables (k_1,σ) and (k_2,ρ); (iv) the extra term must have *no singularities* in any of the variables k_1^2, k_2^2, $k_1\cdot k_2$ and m_0, since the discontinuities of the triangle diagram across its singularities involve no linear divergences and hence are unambiguously contained in Rosenberg's expression $R_{\sigma\rho\mu}$.

(5) The statement in Ref. 20, that the simultaneous presence of isoscalar and isovector vector mesons affects the $\pi^0\rightarrow 2\gamma$ prediction, is not correct. There will, of course, be an extra term of the form

$$\partial B^\xi(I=1)/\partial x_\sigma \partial B^\tau(I=0)/\partial x_\rho \epsilon_{\xi\sigma\tau\rho}$$

in the PCAC equation. However, the matrix element of this term relevant to the $\pi^0\rightarrow 2\gamma$ low-energy theorem, when expressed in terms of Fourier transforms of the vector-meson fields, is proportional to

$$\int d^4k\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|B_{k+k_1+k_2}{}^\xi(I=1)B_{-k}{}^\tau(I=0)|0\rangle \times(k_1+k_2)^\sigma k^\tau\epsilon_{\xi\sigma\tau\rho}.$$

Because of photon gauge invariance, the matrix element

$$\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|B_{k+k_1+k_2}{}^\xi(I=1)B_{-k}{}^\tau(I=0)|0\rangle$$

is proportional to k_1k_2, and so the two-vector meson term is of order $k_1k_2(k_1+k_2)$. Since the low-energy theorem involves only terms of order k_1k_2, the two-vector meson contribution is of higher order and does *not* affect our result. This also means that the extra terms in the PCAC equation proposed recently by R. Arnowitt, M. H. Friedman, and P. Nath, Phys. Letters **27B**, 657 (1968), do not in fact lead to a non-null PCAC prediction for $\pi^0\rightarrow 2\gamma$.

ACKNOWLEDGMENTS

I am grateful to Professor S. B. Treiman for many interesting discussions. I wish to thank Dr. R. J. Eden and Professor A. B. Pippard for the hospitality of the Cavendish high energy group and Clare Hall, where the work reported here was begun. The calculations were completed and the manuscript was written during the 1968 National Accelerator Laboratory summer study at Aspen. I wish to thank Dr. Bell and Dr. Jackiw for an interesting correspondence about their results.

APPENDIX

We discuss here the following questions raised both by the recent paper of Bell and Jackiw and in conversations with Professor S. Coleman: (1) Is the expression $R_{\sigma\rho\mu}$ [see Eq. (17)] which we have used for the triangle graph unique, or is it possible to redefine $R_{\sigma\rho\mu}$ by a subtraction in such a way as to eliminate the anomalies discussed in the text? (2) What is the connection between our results and the σ-model discussion of Bell and Jackiw, and between our results and the physical $\pi^0\rightarrow 2\gamma$ and $\eta\rightarrow 2\gamma$ decays?

A. Uniqueness of the Triangle Graph

The expression for $R_{\sigma\rho\mu}$ in Eq. (17) is obtained from Eq. (16) by the regulator technique of subtracting from

[12] Because of an implicit photon field dependence of $j_0^5(x)$ implied by Eq. (30), $\bar{Q}^5$ *does* commute with all the photon field variables. The details of showing this are complicated, and will be given elsewhere.

[13] J. S. Bell and R. Jackiw (unpublished).

Eq. (16) a loop with m_0 replaced by M, performing the r integration, and then letting $M\to\infty$. Clearly, any *mass-independent* terms in Eq. (16) will be lost in this process. That a mass-independent term is present can be seen from the fact that when we make the change of integration variable $r\to r+ak_1+bk_2$ in Eq. (16), the result is not left invariant, but rather is changed by multiples of $\epsilon_{\tau\sigma\rho\mu}k_1^{\tau}$ and $\epsilon_{\tau\sigma\rho\mu}k_2^{\tau}$. If we are careful to preserve symmetry with respect to the photon variables, the change will be proportional to $\epsilon_{\tau\sigma\rho\mu}(k_1-k_2)^{\tau}$. The noninvariance of the triangle graph under changes of integration variable is of course just a result of the linear divergence in Eq. (16), and means that in a nonregulator calculation the results obtained for the triangle graph will depend on how the external momenta k_1 and k_2 are taken to run through the internal lines. We may express this ambiguity formally by writing that the general expression for the triangle graph is

$$R_{\sigma\rho\mu}[\zeta]=R_{\sigma\rho\mu}+\zeta\epsilon_{\tau\sigma\rho\mu}(k_1-k_2)^{\tau}, \qquad (A1)$$

with $R_{\sigma\rho\mu}$ the regulator value in Eq. (17).

We easily find the following properties of $R_{\sigma\rho\mu}[\zeta]$:

(i) vector index divergence:

$$\begin{aligned} k_1^{\sigma}R_{\sigma\rho\mu}[\zeta]&=-\zeta k_1^{\sigma}k_2^{\tau}\epsilon_{\tau\sigma\rho\mu},\\ k_2^{\rho}R_{\sigma\rho\mu}[\zeta]&=\zeta k_2^{\rho}k_1^{\tau}\epsilon_{\tau\sigma\rho\mu}; \end{aligned} \qquad (A2)$$

(ii) axial-vector index divergence:

$$\begin{aligned} &-(k_1+k_2)^{\mu}R_{\sigma\rho\mu}[\zeta]\\ &\qquad=2m_0R_{\sigma\rho}+(8\pi^2-2\zeta)k_1^{\xi}k_2^{\tau}\epsilon_{\xi\tau\sigma\rho}; \end{aligned} \qquad (A3)$$

(iii) asymptotic behavior: Writing $k_1=\xi q$, $k_2=-\xi q+p'-p$, as $\xi\to\infty$

$$R_{\sigma\rho\mu}[\zeta]\to-\xi(8\pi^2-2\zeta)q^{\tau}\epsilon_{\tau\sigma\rho\mu}; \qquad (A4)$$

(iv) axial-vector meson to two-photon matrix element: If $l\cdot(k_1+k_2)=\epsilon_1\cdot k_1=\epsilon_2\cdot k_2=k_1^2=k_2^2=0$, then[3]

$$l^{\mu}\epsilon_1^{\sigma}\epsilon_2^{\rho}R_{\sigma\rho\mu}[\zeta]=\zeta l^{\mu}\epsilon_1^{\sigma}\epsilon_2^{\rho}(k_1-k_2)^{\tau}\epsilon_{\tau\sigma\rho\mu}; \qquad (A5)$$

(v) large m_0 behavior:

$$\lim_{m_0\to\infty}R_{\sigma\rho\mu}[\zeta]=\zeta\epsilon_{\tau\sigma\rho\mu}(k_1-k_2)^{\tau}. \qquad (A6)$$

Referring first to Eqs. (A2)–(A4), we see that when $\zeta=0$, which is the case discussed in the text, the triangle graph is gauge-invariant with respect to the photon indices but has an anomalous axial-vector Ward identity and anomalous asymptotic behavior. By contrast, when $\zeta=4\pi^2$ there is no longer gauge invariance with respect to the photon indices, but the axial-vector Ward identity and the asymptotic behavior as $\xi\to\infty$ are normal. Since the formal proof of gauge invariance for the triangle graph suffers from the same difficulties as does the formal proof of the axial-vector Ward identity, there is no *a priori* reason to demand gauge invariance with respect to the photon indices as opposed to a normal axial-vector Ward identity, or, for that matter, to demand either. In other words, as long as we consider only the divergence properties of $R_{\sigma\rho\mu}[\zeta]$, there is no requirement fixing ζ.

There are, however, two additional restrictions on $R_{\sigma\rho\mu}$ which force us to choose $\zeta=0$. First of all, we recall[14] that two real photons can never be in a state with total angular momentum 1, which means that the matrix element for an axial-vector meson to decay into two photons must vanish. In order for our triangle graph to satisfy this requirement, we must have $l^{\mu}\epsilon_1^{\sigma}\epsilon_2^{\rho}R_{\sigma\rho\mu}[\zeta]=0$ when l is an axial-vector meson polarization vector satisfying $l\cdot(k_1+k_2)=0$ and when the photon variables satisfy $\epsilon_1\cdot k_1=\epsilon_2\cdot k_2=k_1^2=k_2^2=0$. Referring to Eq. (A5), we see that this requirement forces us to choose $\zeta=0$. [To check that, even with the constraints on l, ϵ, etc., the expression $l^{\mu}\epsilon_1^{\sigma}\epsilon_2^{\rho}(k_1-k_2)^{\tau}\epsilon_{\tau\sigma\rho\mu}$ is in general nonvanishing, choose $k_1=(-1,1,0,0)$, $\epsilon_1=(0,0,1,0)$, $k_2=(-2,0,2,0)$, $\epsilon_2=(0,1,0,0)$, $k_1+k_2=(-3,1,2,0)$, $l=(0,0,0,1)$, $k_1-k_2=(1,1,-2,0)$.] Secondly, it is physically unreasonable that a loop diagram such as our triangle graph should influence *low-energy phenomena* in the limit as the mass of the loop fermion becomes infinite. In other words, we expect

$$\lim_{m_0\to\infty}R_{\sigma\rho\mu}[\zeta]=0,\quad k_1,\,k_2 \text{ fixed} \qquad (A7)$$

which according to Eq. (A6) again requires $\zeta=0$. Thus, *there are strong physical restrictions which uniquely select the regulator value for the triangle graph*; in particular, it is not permissible to make the choice $\zeta=4\pi^2$ which eliminates the anomalies discussed in the text.

B. Connection with Bell and Jackiw and with $\pi^0\to 2\gamma$ and $\eta\to 2\gamma$ Decay

In a recent paper, Bell and Jackiw discuss $\pi^0\to 2\gamma$ in the σ model; they find and attempt to resolve a paradox arising from the presence of triangle diagrams. We briefly summarize their work, and then discuss our own interpretation of the paradox, which differs from theirs.[15] Bell and Jackiw use a truncated version of the σ model, in which the charged pion and the neutron fields are omitted. Letting ψ, ϕ, and σ be, respectively, the fields of the proton, the neutral pion, and the scalar meson, the Lagrangian density is[8]

$$\begin{aligned} \mathcal{L}=\bar\psi[i\gamma\cdot\square-m_0+g_0(\sigma+i\phi\gamma_5)]\psi+\tfrac12[(\partial\phi)^2+(\partial\sigma)^2]\\ -\tfrac12\mu_0^2\phi^2-\tfrac12(\mu_0^2+2\lambda_0/f_0^2)\sigma^2-\lambda_0[(\phi^2+\sigma^2)^2\\ -2f_0^{-1}\sigma(\phi^2+\sigma^2)]-\tfrac14F_{\mu\nu}F^{\mu\nu}-e_0\bar\psi\gamma_\mu\psi A^{\mu}, \end{aligned} \qquad (A8)$$

with the coupling constant f_0 given by

$$f_0=g_0/(2m_0). \qquad (A9)$$

[14] C. N. Yang, Phys. Rev. 77, 242 (1950).

[15] Our results do not contradict those of Bell and Jackiw, but rather complement them. The main point of Bell and Jackiw is that the σ model interpreted in the conventional way, does not satisfy the requirements of PCAC. Bell and Jackiw modify the σ model in such a way as to restore PCAC. We, on the other hand, stay within the conventional σ model, and try to systematize and exploit the PCAC breakdown.

The axial-vector current is

$$j_\mu{}^5(x)=\bar{\psi}(x)\gamma_\mu\gamma_5\psi(x)+2\left[\sigma(x)\frac{\partial}{\partial x^\mu}\phi(x)-\phi(x)\frac{\partial}{\partial x^\mu}\sigma(x)\right]-f_0{}^{-1}\frac{\partial}{\partial x^\mu}\phi(x), \quad \text{(A10)}$$

and the divergence of the axial-vector current, as calculated by *formal use of the equations of motion*, is

$$\frac{\partial}{\partial x_\mu}j_\mu{}^5(x)=\frac{\mu_0{}^2}{f_0}\phi(x). \quad \text{(A11)}$$

This is, of course, the usual operator PCAC equation.

The paradox noted by Bell and Jackiw is obtained by applying Eq. (A11) to the calculation of $\pi^0\rightarrow 2\gamma$ decay. Let us concentrate first on the left-hand side of Eq. (A11). The matrix element $\mathfrak{M}_\mu$ of the axial-vector current between the vacuum and a state with two photons has the following general structure, imposed by the requirements of Lorentz invariance, gauge invariance, and Bose statistics [cf. Eq. (17)]:

$$\begin{aligned}
&\mathfrak{M}_\mu=\epsilon_1{}^\sigma\epsilon_2{}^\rho S_{\sigma\rho\mu}(k_1,k_2)\,,\\
&S_{\sigma\rho\mu}(k_1,k_2)=C_1k_1{}^\tau\epsilon_{\tau\sigma\rho\mu}+C_2k_2{}^\tau\epsilon_{\tau\sigma\rho\mu}+C_3k_{1\rho}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\sigma\mu}\\
&\qquad+C_4k_{2\rho}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\sigma\mu}+C_5k_{1\sigma}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\rho\mu}\\
&\qquad+C_6k_{2\sigma}k_1{}^\xi k_2{}^\tau\epsilon_{\xi\tau\rho\mu}\,,\\
&C_1=k_1\cdot k_2C_3+k_2{}^2C_4\,,\\
&C_2=k_1{}^2C_5+k_1\cdot k_2C_6\,,\\
&C_3(k_1,k_2)=-C_6(k_2,k_1)\,,\\
&C_4(k_1,k_2)=-C_5(k_2,k_1)\,.
\end{aligned} \quad \text{(A12)}$$

As in Eq. (17), k_1 and k_2 denote the photon four-momenta. The matrix element of the divergence of the axial-vector current is proportional to $(k_1+k_2)^\mu\mathfrak{M}_\mu$, and a straightforward algebraic rearrangement[3] using Eq. (A12) shows that

$$(k_1+k_2)^\mu\epsilon_1{}^\sigma\epsilon_2{}^\rho S_{\sigma\rho\mu}(k_1,k_2)\,|_{k_1{}^2=k_2{}^2=0}=\tfrac{1}{2}(C_3-C_6)(k_1+k_2)^2k_1{}^\xi k_2{}^\tau\epsilon_1{}^\sigma\epsilon_2{}^\rho\epsilon_{\xi\tau\sigma\rho}. \quad \text{(A13)}$$

Thus, if we write the matrix element for $\pi^0\rightarrow 2\gamma$ in the form

$$\mathfrak{M}(\pi^0\rightarrow 2\gamma)=k_1{}^\xi k_2{}^\tau\epsilon_1{}^\sigma\epsilon_2{}^\rho\epsilon_{\xi\tau\sigma\rho}F\,, \quad \text{(A14)}$$

then Eqs. (A11) and (A13) tell us that in the σ model (or any other PCAC model), F vanishes when the pion mass $(k_1+k_2)^2$ is extrapolated to zero. This statement, of course, must hold in each order of perturbation theory. So let us check by calculating $\mathfrak{M}(\pi^0\rightarrow 2\gamma)$ directly in the σ model in lowest-order perturbation theory, where the only diagram which contributes is the pseudoscalar coupling triangle diagram (i.e., Fig. 2 with $\gamma_\mu\gamma_5$ replaced by the pion-nucleon coupling $ig_0\gamma_5$). We find, comparing with Eqs. (19) and (20), that

$$\begin{aligned}
\mathfrak{M}(\pi^0\rightarrow 2\gamma)_{\text{lowest order}}&=\frac{-ie_0{}^2}{(2\pi)^4}ig_0\epsilon_1{}^\sigma\epsilon_2{}^\rho R_{\sigma\rho}\\
&=k_1{}^\xi k_2{}^\tau\epsilon_1{}^\sigma\epsilon_2{}^\rho\epsilon_{\xi\tau\sigma\rho}(2\alpha_0/\pi)g_0m_0I_{00}(k_1,k_2)\,,
\end{aligned} \quad \text{(A15)}$$

so that

$$F_{\text{lowest order}}=\frac{2\alpha_0}{\pi}g_0m_0I_{00}(k_1,k_2)\,|_{k_1{}^2=k_2{}^2=0}. \quad \text{(A16)}$$

Setting $(k_1+k_2)^2=0$ then gives

$$F_{\text{lowest order}}\,|_{(k_1+k_2)^2=0}=-\frac{\alpha_0}{\pi}\frac{g_0}{m_0}\,, \quad \text{(A17)}$$

which *does not vanish*, contradicting the conclusion obtained indirectly from PCAC. The nonzero value of Eq. (A17) is the paradox of Bell and Jackiw.

Bell and Jackiw attempt to circumvent this contradiction by introducing a regulator nucleon field ψ_1 which is quantized with commutators rather than anticommutators. The coupling of the regulator field to the mesons is described by the interaction Lagrangian density

$$\bar{\psi}_1g_1(\sigma+i\phi\gamma_5)\psi_1\,; \quad \text{(A18)}$$

to maintain the PCAC equation the regulator coupling and mass must satisfy the relation

$$g_1/m_1=g_0/m_0. \quad \text{(A19)}$$

Thus, as the regulator mass approaches infinity, the regulator coupling to the mesons becomes infinite as well. As a consequence, even in the limit of infinite regulator mass the regulator field triangle diagram makes a contribution to the amplitude for $\pi^0\rightarrow 2\gamma$ decay,

$$F_{\text{regulator triangle diagram}}\xrightarrow[m_1\rightarrow\infty]{}\frac{\alpha_0}{\pi}\frac{g_1}{m_1}=\frac{\alpha_0}{\pi}\frac{g_0}{m_0}. \quad \text{(A20)}$$

The total amplitude is the sum of Eqs. (A16) and (A20), and *does* vanish at $(k_1+k_2)^2=0$, in accord with the PCAC prediction.

Unfortunately, however, the regulator procedure of Bell and Jackiw leads to grave difficulties when we turn to purely strong interaction phenomena. Let us, in particular, consider the regulator loop contribution to the scattering of $2n$ σ particles. In the limit of large regulator mass, this loop is proportional to

$$g_1{}^{2n}\int d^4r\,\mathrm{Tr}\left\{\left[\frac{1}{r-m_1}\right]^{2n}\right\}\propto m_1{}^4\left(\frac{g_1}{m_1}\right)^{2n}, \quad \text{(A21)}$$

and thus, on account of Eq. (A19), becomes infinite as $m_1\rightarrow\infty$. This means that the regulator procedure of Bell and Jackiw introduces unrenormalizable infinities into the strong interactions in the σ model, and therefore is not satisfactory.

We now suggest a different resolution of the paradox, utilizing the ideas developed in the text.[15] As we saw, when triangle graphs are present we cannot naively use the equations of motion to calculate the divergence of the axial-vector current. Rather, we must infer the correct divergence equation from perturbation theory, which tells us that the extra term of Eq. (30) is present. In the σ model, the effect of this extra term is to replace Eq. (A11) by

$$\frac{\partial}{\partial x_\mu} j_\mu{}^5(x) = \frac{\mu_0{}^2}{f_0}\phi(x) + \frac{\alpha_0}{4\pi} F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}. \tag{A22}$$

In other words, *the PCAC equation must be modified in the presence of electromagnetic interactions.* As a result, the argument leading to the conclusion that F vanishes at $(k_1+k_2)^2=0$ must be modified. As before, we conclude that the matrix element of the left-hand side of Eq. (A22) between vacuum and two photons vanishes at $(k_1+k_2)^2=0$. But instead of implying that $\mathfrak{M}(\pi^0\to 2\gamma)$ vanishes, this now tells us that

$$\begin{aligned}\mathfrak{M}(\pi^0\to 2\gamma) &= Z_3{}^{-1/2}\times\text{matrix element of } (\mu^2\phi)\\ &= -\mu^2(f_0/\mu_0{}^2)Z_3{}^{-1/2}\times\text{matrix element}\\ &\qquad\text{of } [(\alpha_0/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}]\\ &= \frac{\mu^2}{\mu_0{}^2}Z_3{}^{-1/2}\left(-\frac{\alpha}{\pi}\frac{g_0}{m_0}\right)k_1{}^\xi k_2{}^\tau \epsilon_1{}^\sigma\epsilon_2{}^\rho\epsilon_{\xi\tau\sigma\rho};\end{aligned} \tag{A23}$$

in other words,

$$F|_{(k_1+k_2)^2=0} = \frac{\mu^2}{\mu_0{}^2}Z_3{}^{-1/2}\left(-\frac{\alpha}{\pi}\frac{g_0}{m_0}\right). \tag{A24}$$

[In Eqs. (A23) and (A24), Z_3 is the π^0 wave-function renormalization constant.] To lowest order in perturbation theory, Eq. (A24) agrees with Eq. (A17), so our modified PCAC equation leads to no paradox. In addition, Eq. (A22) yields a bonus: From the derivation of Eq. (A24) it is clear that Eq. (A24) is not just a lowest-order perturbation theory result, but in fact is an *exact* statement in the σ model. We can reexpress Eq. (A24) in terms of physical quantities using the equation[16]

$$\frac{g_0}{m_0}\frac{\mu^2}{\mu_0{}^2}Z_3{}^{-1/2} = \frac{g_r(0)}{m_N}\frac{1}{g_A}, \tag{A25}$$

where m_N, $g_r(0)$, g_A are, respectively, the renormalized nucleon mass, the renormalized pion-nucleon coupling constant (evaluated at pion mass zero), and the nucleon axial-vector coupling constant in the σ model. Thus Eq. (A24) becomes

$$F|_{(k_1+k_2)^2=0} = -\frac{\alpha}{\pi}\frac{g_r(0)}{m_N g_A}. \tag{A26}$$

[16] M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960).

Let us now make the standard PCAC assumption that F is slowly varying as the pion mass $(k_1+k_2)^2$ is varied from μ^2 to 0, so that we can use Eq. (A26) for the physical π^0-decay matrix element. We also replace $g_r(0)$ by the on-shell coupling constant g_r. Using the physical values for μ, m_N, g_r, g_A,[17] we find for the pion lifetime

$$\tau^{-1} = (\mu^3/64\pi)F^2 = 9.7 \text{ eV}, \tag{A27}$$

in good agreement with the experimental value[18]

$$\begin{aligned}\tau_{\text{expt}}{}^{-1} &= (1.12\pm 0.22)\times 10^{16} \text{ sec}^{-1}\\ &= (7.37\pm 1.5) \text{ eV}.\end{aligned} \tag{A28}$$

So we see that the σ model, as interpreted with Eq. (A22), gives a reasonable account of $\pi^0\to 2\gamma$ decay.[19] This also makes it clear that the use of regulators to cancel away the triangle graph contribution to F up to terms of order $\mu^2/m_N{}^2$ will tend to give much too small a value for the $\pi^0\to 2\gamma$ matrix element.

The above ideas are readily extended to other field theoretical models, and hopefully, to the physical axial-vector current as well. Let $\mathfrak{F}_3{}^{5\lambda}$ be the third component of the axial-vector octet. (It corresponds to $\frac{1}{2}j^{5\lambda}$ in the model discussed above.) Let us suppose that the world is really described by a field theory, and that there are only spin-0 or spin-$\frac{1}{2}$ elementary fields.[20] We then make the following two assumptions:

(i) The usual PCAC equation,

$$\frac{\partial}{\partial x^\lambda}\mathfrak{F}_3{}^{5\lambda} = C_\pi \mu^2 \phi_{\pi^0}, \quad C_\pi = m_N g_A/g_r(0), \tag{A29}$$

[17] We take $g_r\approx 13.4$, $g_A\approx 1.18$. If we used $g_A\approx 1.24$, then we would get $\tau^{-1}=8.9$ eV. We can also evaluate Eq. (A26) by using the relation $g_r(0)/(m_N g_A)=\sqrt{2}\mu_+{}^2/f_\pi$, with f_π the charged-pion decay amplitude and μ_+ the charged-pion mass. (See S. L. Adler and R. F. Dashen, Ref. 11, pp. 41–45.) This gives $F|_{(k_1+k_2)^2=0} = -(\alpha/\pi)\sqrt{2}\mu_+{}^2/f_\pi$. Using the experimental value $f_\pi\approx 0.96\ \mu_+{}^3$, we find from Eq. (A27) that $\tau^{-1}=7.4$ eV.

[18] A. H. Rosenfeld *et al.*, Rev. Mod. Phys. 40, 77 (1968).

[19] Comparing Eqs. (A26) and (A17), we see that apart from a factor of $g_A{}^{-2}$, our PCAC expression for the π^0 lifetime is the same as the expression obtained from the pseudoscalar coupling triangle graph if one uses the physical nucleon mass and pion-nucleon coupling rather than the bare mass and coupling appearing in Eq. (A17). That the triangle graph, evaluated using physical quantities, gives a good value for $\pi^0\to 2\gamma$ decay has been noted by J. Steinberger, Phys. Rev. 76, 1180 (1949); and J. Steinberger (private communication).

[20] This assumption is not strictly necessary for the calculation of the $\pi^0\to 2\gamma$ rate. If there is also a single elementary neutral vector-meson field B^λ, then there will be an additional term in Eq. (A30) proportional to $F^{\xi\sigma}\partial B^\tau/\partial x_\rho\epsilon_{\xi\sigma\tau\rho}$. However, because the gauge-invariant coupling of a massive vector boson to a physical photon vanishes [G. T. Feldman and P. T. Matthews, Phys. Rev. 132, 823 (1963)], this term makes no contribution to the physical $\pi^0\to 2\gamma$ decay. In general, there will be no change in the $\pi^0\to 2\gamma$ prediction if *only* isoscalar vector mesons or *only* isovector vector mesons are present. If *both* isoscalar and isovector vector mesons are present, there will be additional terms like $\partial B^\xi(I=1)/\partial x_\sigma\, \partial B^\tau(I=0)/\partial x_\rho\epsilon_{\xi\sigma\tau\rho}$, which do affect the $\pi^0\to 2\gamma$ prediction.

should, on account of triangle graphs, be replaced by

$$\frac{\partial}{\partial x^\lambda}\mathfrak{F}_3^{5\lambda}=C_\pi\mu^2\phi_{\pi^0}+S\frac{\alpha_0}{4\pi}F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}, \quad \text{(A30)}$$

with S a constant.[21]

(ii) If $\mathfrak{F}_3^{5\lambda}$ is expressed in terms of the elementary fields by

$$\mathfrak{F}_3^{5\lambda}=\sum_j g_j\bar{\psi}_j\gamma^\lambda\gamma^5\psi_j+\text{meson terms}, \quad \text{(A31)}$$

then S is given by

$$S=\sum_j g_jQ_j^2, \quad \text{(A32)}$$

where the charge of the jth fermion is Q_je_0. Equation (A32) means that we count only triangle graphs of the elementary fermions, but do not include triangles involving nonelementary bound states. It may be possible to decide in model calculations whether this rule, which we conjecture, is really correct.

Using Eq. (A30) to calculate the $\pi^0\to 2\gamma$ matrix element then gives

$$F\approx-(\alpha/\pi)2S(g_r/m_Ng_A). \quad \text{(A33)}$$

The experimentally measured π^0 lifetime corresponds[22] to $|S|=0.44$; for comparison, S in the σ model is $\frac{1}{2}1^2-\frac{1}{2}0^2=\frac{1}{2}$, while S in the quark model is $\frac{1}{2}(\frac{2}{3})^2-\frac{1}{2}(-\frac{1}{3})^2=\frac{1}{6}$. More generally, in any triplet model in which the electromagnetic current is a U-spin singlet, the triplet charges will be $(Q_p,Q_n,Q_\lambda)=(Q,Q-1,Q-1)$ and we have $S=\frac{1}{2}Q^2-\frac{1}{2}(Q-1)^2=Q-\frac{1}{2}$. That is, in triplet models we have $S=\langle Q\rangle_{av}$, where $\langle Q\rangle_{av}$ is the average charge of the triplet particles taking part in both the $\Delta S=0$ weak $V-A$ current and the $|\Delta S|=1$ weak $V-A$ current. This means that the condition $\langle Q\rangle_{av}=-\frac{1}{2}$, necessary[23] for the radiative corrections to the $\Delta S=0$ and $|\Delta S|=1$ weak currents to be finite, also predicts a $\pi^0\to 2\gamma$ rate in good accord with experiment.[24]

The two-photon decay $\eta\to 2\gamma$ can be treated in a similar manner. The analog of Eq. (A30) for $\mathfrak{F}_8^{5\lambda}$ is

$$\frac{\partial}{\partial x^\lambda}\mathfrak{F}_8^{5\lambda}=C_\eta\mu_\eta^2\phi_\eta+\frac{1}{\sqrt{3}}S\frac{\alpha_0}{4\pi}F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}, \quad \text{(A34)}$$

where S is the same constant as in Eq. (A30) and where the factor $3^{-1/2}$ appears because the electromagnetic current is a U-spin singlet.[25] If there were no $\eta - X^0$ mixing, then ϕ_η would be the η field; in the presence of mixing, ϕ_η would be a mixture of the η and X^0 fields. In the SU_3 limit, one has, of course, $C_\eta=C_\pi$. To get a prediction for the $\eta\to 2\gamma$ rate from Eq. (A34), we sandwich Eq. (A34) between the η state and a two-photon state and make the following three approximations: (i) We neglect $\eta-X^0$ mixing; (ii) we take $C_\eta=C_\pi$; (iii) we ngelect the left-hand side of Eq. (A34), which makes a contribution of order μ_η^2 [equivalently, we assume that the *exact* prediction $F_\eta(\mu_\eta^2=0)=-(\alpha/\pi)\times(2S/\sqrt{3})(1/C_\eta)$ can be smoothly extrapolated from $\mu_\eta^2=0$ to the physical η mass]. These approximations give the standard SU_3 prediction[26]

$$\Gamma(\eta\to 2\gamma)=\tfrac{1}{3}(\mu_\eta/\mu)^3\Gamma(\pi^0\to 2\gamma)=(165\pm34)\text{ eV}, \quad \text{(A35)}$$

about a factor of 8 smaller than the experimental value of

$$\Gamma(\eta\to 2\gamma)=(1210\pm260)\text{ eV}. \quad \text{(A36)}$$

In view of the approximations made, the discrepancy is not too disturbing; in particular, the terms of order μ_η^2 are by no means negligible, and could easily make a contribution to the $\eta\to 2\gamma$ matrix element as important as the $S/\sqrt{3}$ term which we have retained.[27]

---

[21] In Eq. (A30), ϕ_{π^0} does not necessarily mean a canonical pion field, but only a suitable interpolating field for the pion. For example, in the quark model, ϕ_{π^0} would be proportional to $\bar{\psi}\gamma_5\tau_3\psi$. The separation of $\partial_\lambda\mathfrak{F}_3^{5\lambda}$ into two terms in Eq. (A30) is made unique by the requirement that ϕ_{π^0} and the photon field be ***dynamically independent***, in the sense that $[\phi_{\pi^0},A_\lambda]=[\phi_{\pi^0},\dot{A}_\lambda]=0$ at equal times.

[22] If we use instead of Eq. (A33) the formula $F\approx-(\alpha/\pi)(2S)\times(\sqrt{2}\mu_+^2/f_\pi)$, as in Ref. 17, then the experimentally measured π^0 lifetime gives $|S|=0.50$.

[23] N. Cabibbo, L. Maiani, and G. Preparata, Phys. Letters **25B**, 132 (1967); K. Johnson, F. Low, and H. Suura, Phys. Rev. Letters **18**, 224 (1967).

[24] This result was noted previously, in the context of the vector dominance model, by N. Cabibbo, L. Maiani, and G. Preparata, Phys. Letters **25B**, 31 (1967).

[25] The correctness of the factor $1/\sqrt{3}$ is easily verified in the triplet model.

[26] The factor $(\mu_\eta/\mu)^3$ comes from phase space.

[27] We discuss briefly two other electromagnetic decays to which current algebra methods have been applied: $\omega\to\pi^0\gamma$ and $\eta\to 3\pi$. In the case of $\omega\to\pi^0\gamma$ it has been argued by D. G. Sutherland [Nucl. Phys. **B2**, 433 (1967)] that the usual PCAC equation [Eq. (A11)] implies vanishing of the decay amplitude at zero π^0 four-momentum. This conclusion, however, is erroneous, and results from the use by Sutherland of an insufficiently general form for the axial-vector-current-vector-meson-photon vertex. The most general such vertex is given by Eq. (A12); an examination of the reasoning leading to Eq. (A13) shows that Eq. (A13) is valid *only* when $k_1^2=k_2^2=0$. When one of the vectors is massive, as in the case of ω decay, we find instead that

$$(k_1+k_2)^\mu\epsilon_1^\sigma\epsilon_2^\rho S_{\sigma\rho\mu}(k_1,k_2)|_{(k_1+k_2)^2=k_1^2=0}=[C_4+C_6-\tfrac{1}{2}(C_5+C_6)]k_2^2k_1^\xi k_2^\tau\epsilon_1^\sigma\epsilon_2^\rho\epsilon_{\xi\tau\sigma\rho}\neq 0,$$

contradicting Sutherland's conclusion. This equation also means that our modified PCAC prediction for $\pi^0\to 2\gamma$ will be altered when one of the photons is virtual, as is the case in the Primakoff effect.

In the decay $\eta\to 3\pi$, the only point which we wish to make is that the triangle graphs which we have considered (involving either photons or strongly interacting vector mesons) cannot alter the usual PCAC predictions. The reason is the presence in all matrix elements coming from our extra term of the factor $k_1^\xi k_2^\tau\times\epsilon_1^\sigma\epsilon_2^\rho\epsilon_{\xi\tau\sigma\rho}$, which vanishes at zero four-momentum for the axial-vector vertex. (In the $\pi^0\to 2\gamma$ case we were always talking about the matrix element left after removal of this factor.)

π^0 DECAY

Stephen L. Adler

Institute for Advanced Study

Princeton, New Jersey

I wish to describe some recent theoretical work on $\pi^0 \to 2\gamma$ decay, which helps to resolve puzzling questions which have arisen over the years, and which may shed light on the nature of possible fundamental constituents of matter, such as quarks. Purely kinematic considerations tell us that the matrix element and the decay rate for this process are

$$\mathcal{M}(\pi^0 \to 2\gamma) = k_1{}^{\xi} k_2{}^{\tau} \epsilon_1{}^{*\sigma} \epsilon_2{}^{*\rho} \epsilon_{\xi\tau\sigma\rho} F ,$$
$$\tau^{-1} = (\mu^3/64\pi)F^2 , \qquad (1)$$

with (k_1, ϵ_1), (k_2, ϵ_2) the momentum and polarization four-vectors of the two photons, μ the pion mass, and F an intrinsic coupling constant. The job for the theorist, of course, is to try to calculate F. An important step towards this goal was taken in 1949 by Steinberger,[1] who considered a model in which the π^0 dissociates (via pseudoscalar coupling) into a proton-antiproton pair, which emit the two photons and then annihilate. In lowest order perturbation theory there are only two Feynman diagrams, the triangle diagram in Fig. 1a and the corresponding diagram with the two photons interchanged. Although this diagram appears to be linearly divergent, the presence of the γ_5 in the Fermion trace causes all divergent terms to vanish identically, and a straightforward calculation gives (neglecting small terms of order $\mu^2/m_N{}^2$)

$$F \approx -\frac{\alpha}{\pi}\frac{g_r}{m_N} , \qquad (2)$$

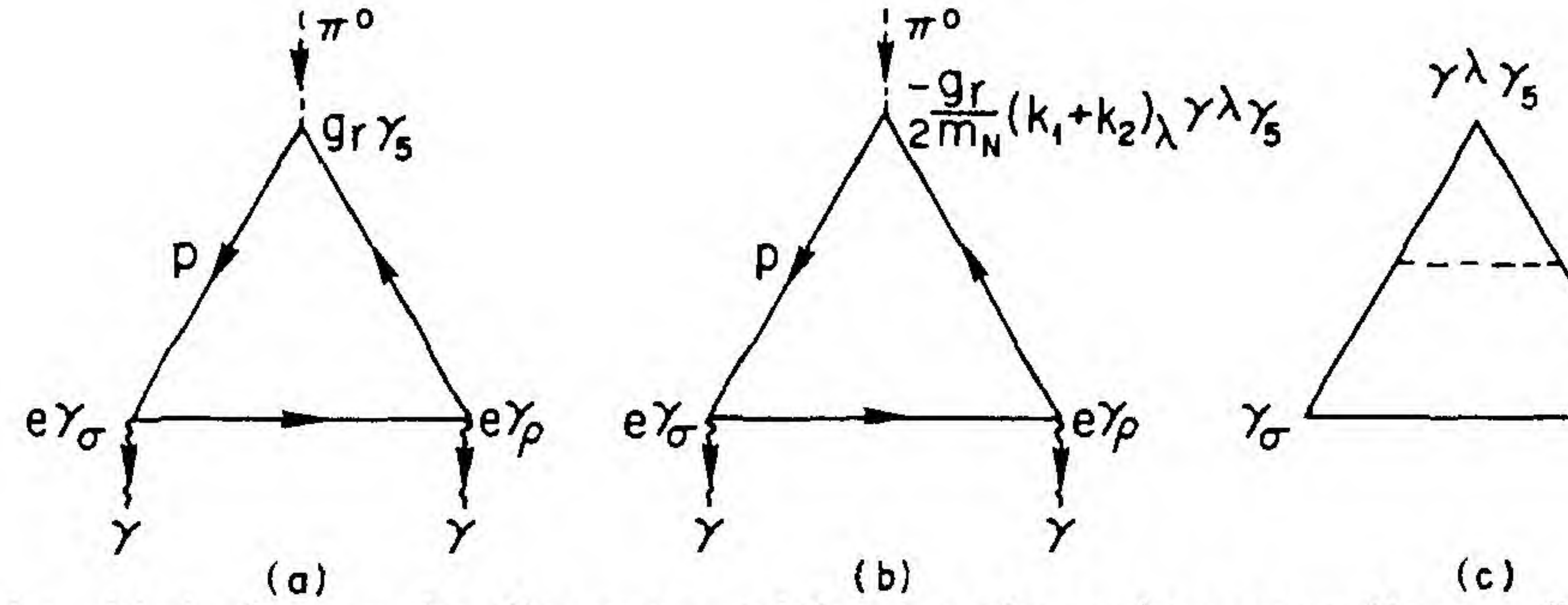


Fig. 1(a) Triangle diagram with pseudoscalar coupling. (b) Triangle diagram with pseudovector coupling. (c) A virtual meson correction to the triangle diagram.

α = fine structure constant = $e^2/4\pi \approx 1/137$,
g_r = pion-nucleon coupling constant ≈ 13.6 ,
m_N = nucleon mass .

Substituting Eq. (2) into Eq. (1), one finds a decay rate $\tau^{-1} \approx 14$ eV, in fairly good agreement with the experimental rate $\tau^{-1}_{\text{expt}} = (1.12 \pm 0.22) \times 10^{16}\ \text{sec}^{-1} = (7.37 \pm 1.5)$ eV. That such a naive calculation should work so well is, in fact, rather puzzling, since we know that Eq. (1) is just the first term in a power series in the strong coupling g_r, and there is no obvious reason why one should be able to get away with the neglect of all of the higher terms.

A second puzzle also emerged from Steinberger's calculation. In addition to calculating $\pi^0 \to 2\gamma$ decay using pseudoscalar coupling, Steinberger also repeated the calculation with pseudovector (axial-vector) coupling, by evaluating the diagram shown in Fig. 1b. This diagram is actually linearly divergent, and must be evaluated by regulator techniques to insure photon gauge invariance. On the basis of the pseudoscalar-pseudovector equivalence theorem, one would expect Fig. 1b to give the same rate as Fig. 1a, but actual calculation shows that Fig. 1b gives a $\pi^0 \to 2\gamma$ amplitude F smaller than that of Eq. (2) by a factor $\mu^2/6m_N{}^2$. In the limit of zero pion mass, the pseudovector amplitude F actually vanishes!

During the last ten years, extensive and very successful calculations on soft pion emission have been done using the partially-conserved axial-vector current (PCAC) hypothesis. This hypothesis states that, apart from certain equal-time commutator terms (which do not enter into our problem), soft pions behave as if they were coupled to nucleons by pseudovector, rather than pseudo-

scalar, coupling. When we turn to Steinberger's calculation, PCAC thus leads us to the troublesome conclusion that the answer $F \approx 0$ should be chosen over the numerically reasonable answer for F given by Eq. (2)! This conclusion is independent of the perturbation theory model used by Steinberger, and is easily derived in a completely general fashion.[2] All we need do is to sandwich the PCAC equation

$$\partial_\lambda \mathfrak{F}_3^{5\lambda} = (f_\pi/\sqrt{2})\phi_{\pi^0} , \qquad (3)$$

f_π = charged pion decay amplitude ,

between the two photon state $\langle \gamma(k_1, \epsilon_1)\gamma(k_2, \epsilon_2)|$ and the vacuum $|0\rangle$. The matrix element of the right-hand side of Eq. (3) is proportional to $\mathcal{M}(\pi^0 \to 2\gamma)$, while a purely kinematic analysis of $\langle \gamma(k_1, \epsilon_1)\gamma(k_2, \epsilon_2)|\mathfrak{F}_3^{5\lambda}|0\rangle$ shows that the matrix element of the left-hand side of Eq.(3) is proportional to $k_1{}^\xi k_2{}^\tau \epsilon_1^{*\sigma} \epsilon_2^{*\rho} \epsilon_{\xi\tau\sigma\rho} \times (k_1+k_2)^2$. Thus, in a model-independent way, PCAC predicts $F \propto (k_1+k_2)^2 = \mu^2$, just as was found from the pseudovector triangle graph in perturbation theory.

To summarize, the theory of π^0 decay presents the following three puzzles:

(1) Naive calculation using the lowest order pseudoscalar coupling triangle diagram, and neglecting possible strong interaction renormalization effects, gives a surprisingly good result.

(2) The pseudoscalar-pseudovector equivalence theorem breaks down. Pseudovector coupling predicts that $\pi^0 \to 2\gamma$ decay is strongly suppressed.

(3) PCAC implies, in a model independent way, that $\pi^0 \to 2\gamma$ decay is strongly suppressed.

Recent theoretical work,[3] which I will now briefly describe, has helped to resolve these puzzles. The key observation is that when very singular diagrams (e.g. triangle diagrams) are present, formal field-theory results such as Ward identities, the pseudoscalar-pseudovector equivalence theorem, and the PCAC equation itself, break down. Consider, for example, the pseudoscalar-pseudovector equivalence theorem, which asserts that the diagrams of Figs. 1a and 1b should give identical results. The theorem is formally derived by taking the vacuum to two photon matrix element of the divergence equation

$$(g_r/2im_N)\partial_\lambda j^{5\lambda} = g_r j^5 \;; \qquad j^{5\lambda} = \overline{\psi}\gamma^\lambda\gamma_5\psi \;, \quad j^5 = \overline{\psi}\gamma_5\psi \;. \tag{4}$$

[We can neglect meson terms in Eq. (4) because no virtual mesons appear in Figs. 1a and 1b.] The matrix element of the right-hand side of Eq. (4) corresponds to Fig. 1a, while the matrix element of the left-hand side of Eq. (4) corresponds to Fig. 1b, and Eq. (4) asserts that they should be equal. This formal derivation breaks down because the local product of operators $\overline{\psi}(x)\gamma^\lambda\gamma_5\psi(x)$ is singular, and the naive manipulations of equations of motion which lead to Eq. (4) are incorrect. The correct answer can be obtained by regarding the axial-vector current and the pseudoscalar current as <u>limits</u> of nonlocal, gauge-invariant currents,[4]

$$j^{5\lambda}(x) = \lim_{\epsilon\to 0} \overline{\psi}(x+\epsilon)\gamma^\lambda\gamma_5\psi(x-\epsilon)\exp\left[-ie\int_{x-\epsilon}^{x+\epsilon} d\ell\cdot A(\ell)\right] ,$$

$$j^5(x) = \lim_{\epsilon\to 0} \overline{\psi}(x+\epsilon)\gamma_5\psi(x-\epsilon)\exp\left[-ie\int_{x-\epsilon}^{x+\epsilon} d\ell\cdot A(\ell)\right] , \tag{5}$$

$$A = \text{electromagnetic field,}$$

giving, after a careful calculation

$$\partial_\lambda j^{5\lambda} = 2im_N j^5 + (\alpha/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho} \;, \qquad F^{\xi\sigma} = \partial^\sigma A^\xi - \partial^\xi A^\sigma = \text{electromagnetic field-strength tensor} \;. \tag{6}$$

The matrix element of the extra term in Eq. (6) precisely accounts for the difference between Figs. 1a and 1b as calculated by Steinberger! An alternative procedure[3] is to directly calculate the difference of Figs. 1a and 1b in momentum space. If one freely translates the loop integration variables one "deduces" that the difference is zero, but if one pays careful attention to the fact that in linearly divergent integrals the origin of integration <u>cannot</u> be freely shifted, one reproduces the right-hand side of Eq. (6). We see, then, that the pseudoscalar-pseudovector equivalence theorem breaks down for triangle diagrams because of the presence of singular (linearly divergent) integrals, and the breakdown is compactly summarized by Eq. (6).

Clearly, the phenomenon which modifies Eq. (6) will also affect the PCAC equation, Eq. (3). Let us consider a particular field-theoretic model, the σ-model of Gell-Mann and Lévy. This model

consists of a neutron n and a proton p interacting with the pions (π^+, π^0, π^-) and with a scalar, isoscalar meson σ via $SU_2 \otimes SU_2$ symmetric couplings. In the absence of electromagnetism, Eq. (3) is satisfied as an operator identity, with ϕ_{π^0} the canonical pion field. In the presence of electromagnetism, the singularity triangle diagram changes Eq. (3) to read

$$\partial_\lambda \mathcal{F}_3^{5\lambda} = (f_\pi/\sqrt{2})\phi_{\pi^0} + \tfrac{1}{2}(\alpha/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho} \quad . \tag{7}$$

In other words, the PCAC equation for the neutral pion must be modified in the presence of electromagnetic interactions. Just as we did above, let us take the matrix element of Eq. (7) between the vacuum and the two-photon state. In the soft pion limit, the matrix element of the left-hand side makes no contribution, but instead of deducing that the $\pi^0 \rightarrow 2\gamma$ amplitude F vanishes, we now find that F is proportional to the matrix element of the extra term in Eq. (7),

$$F\bigg|_{(k_1+k_2)^2=0} = -\frac{\alpha}{\pi}\frac{\sqrt{2}\,\mu^2}{f_\pi} \quad . \tag{8}$$

Because Eq. (8) has been derived without recourse to perturbation theory, it is an exact low energy theorem for π^0 decay.[5] Using the experimental value $f_\pi \approx 0.96\,\mu^3$ and substituting Eq. (8) into Eq. (1), we find the decay rate $\tau^{-1} = 7.4$ eV, in excellent agreement with experiment. It is interesting to compare Eq. (8) with Steinberger's lowest order perturbation theory result [Eq. (2)] by using the Goldberger-Treiman relation,

$$\frac{\sqrt{2}\,\mu^2}{f_\pi} \approx \frac{g_r}{m_N g_A} \quad , \tag{9}$$

g_A = nucleon axial-vector coupling constant ≈ 1.22,

to rewrite Eq. (8) in the form

$$F\bigg|_{(k_1+k_2)^2=0} \approx -\frac{\alpha}{\pi}\frac{g_r}{m_N}\frac{1}{g_A} \quad . \tag{10}$$

We see that the effects of higher orders of perturbation theory are entirely contained in the factor g_A^{-1}, which is numerically close to unity; as a result Steinberger's calculation, which neglects the factor g_A, is a fairly good first approximation.

We see, then, that the modified PCAC equation resolves the puzzles noted above. At the same time, however, some new

questions and problems are raised. Let us now consider these problems, as well as some of the experimental implications of Eq. (7).

1. We have just gone through some rather subtle reasoning to avoid the prediction, following from the unmodified PCAC equation, that F is suppressed by a factor $\sim \mu^2/m_N^2$. However, one can always ask how one knows that π^0 decay is not really a suppressed decay. One argument is the theoretical one, that with the extra term PCAC gives a good answer for π^0 decay, which means that without this term, the rate would be much too small to agree with experiment. There is also an interesting experimental test,[6] which strongly suggests that π^0 decay is not suppressed. To see this, let us return to the suppression argument following Eq. (3), in the altered situation in which one of the photons is off-mass-shell, say $k_1^2 \neq 0$. Some simple kinematics shows that the vacuum to two photon matrix element of $\partial_\lambda \mathcal{F}_3^{5\lambda}$ is now proportional to $k_1{}^\xi k_2{}^\tau \epsilon_1^{*\sigma} \epsilon_2^{*\rho} \times \epsilon_{\xi\tau\sigma\rho}[\mu^2+\beta k_1{}^2]$, with β of order unity. We see that while the on-shell part of the amplitude is suppressed by a factor μ^2, the off-shell dependence is not suppressed. Since the off-shell amplitude is measured in the reaction $\pi^0 \to e^+e^-\gamma$, our suppression argument predicts that the $k_1{}^2$ dependence of this process will have the form $1 + (\beta/\mu^2)k_1{}^2$, which has a much larger slope than the form $1 + (\beta/m_\rho{}^2)k_1{}^2$ expected in the absence of suppression of the $\pi^0 \to 2\gamma$ decay. A measurement of this slope has been reported by Devons et al.,[7] who find a matrix element $1 + \underline{a}\, k_1{}^2$, $\underline{a} = (0.01 \pm 0.11)/\mu^2$. Clearly, this is strong evidence against $\pi^0 \to 2\gamma$ suppression.

2. The argument that Eq. (8) is an exact low energy theorem is not as simple as we have made it sound. To be sure that Eq. (8) is exact, we must be sure that strong interaction modifications of the triangle diagram, such as illustrated in Fig. 1c, do not renormalize the extra term in Eq. (7). This can in fact be demonstrated, to any finite order of perturbation theory.[5] The reason that virtual meson corrections do not modify Eq. (7) is that they always involve Fermion loops with more than three vertices (Fig. 1c involves a 5-vertex loop), which satisfy normal Ward identities because they are highly convergent. There is still the possibility that Eq. (7) is modified by nonperturbative effects, such as contributions from triangles involving bound states of the fundamental fields. Our neglect of possible nonperturbative modifications is pure assumption.

3. The spectacularly good agreement of Eq. (8) with experiment is

somewhat fortuitous, both because of the large error in the experimental $\pi^0 \to 2\gamma$ rate and because of the usual 10-20 percent extrapolation error involved in PCAC arguments. For example, use of the Goldberger-Treiman relation to replace Eq. (8) by Eq. (10) alters the theoretical prediction by 20 percent, to $\tau^{-1} = 9.1$ eV.

4. The constant $\frac{1}{2}$ appearing in front of the term $(\alpha/4\pi)F^{\xi\sigma}F^{\tau\rho} \times \epsilon_{\xi\sigma\tau\rho}$ in Eq. (7) arises from our particular choice of fundamental Fermion fields, and differs in different field-theoretic models. Quite generally, if $\mathcal{F}_3^{5\lambda}$ is expressed in terms of fundamental fields by

$$\mathcal{F}_3^{5\lambda} = \sum_j g_j \bar{\psi}_j \gamma^\lambda \gamma_5 \psi_j + \text{meson terms} \quad , \tag{11}$$

then the modified PCAC reads

$$\begin{aligned} \partial_\lambda \mathcal{F}_3^{5\lambda} &= (f_\pi/\sqrt{2})\phi_{\pi^0} + S(\alpha/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho} \quad , \\ S &= \sum_j g_j Q_j^2 \quad , \end{aligned} \tag{12}$$

where the charge of the j^{th} fermion is Q_j. All we are doing, of course, is adding up the contributions of the triangle diagrams involving the various Fermions. The $\pi^0 \to 2\gamma$ low energy theorem derived from Eq. (12) is

$$F \Big|_{(k_1+k_2)^2 = 0} = -\frac{\alpha}{\pi}\frac{\sqrt{2}\,\mu^2}{f_\pi} 2S \approx -\frac{\alpha}{\pi}\frac{g_r}{m_N}\frac{1}{g_A} 2S \quad , \tag{13}$$

which reduces to our previous result when $S = \frac{1}{2}$.

The comparison which we have made above with the experimental π^0 decay rate tells us that $|S| \approx 0.5$, but does not determine the <u>sign</u> of S. However, there are a number of different ways of determining the sign of S, all of which, fortunately, seem to agree! The first method is by analysis of $\pi^+ \to e^+\nu\gamma$ decay, the vector part of which is related by CVC to F and the axial-vector part of which can be estimated by hard pion techniques. Using the experimentally measured vector to axial-vector ratio for this process, Okubo[8] finds that S is positive. A second method is to make use of forward π^0 photoproduction, where one can observe the interference between the Primakoff amplitude (which is proportional to F) and the forward purely strong interaction amplitude. The sign of the latter can be determined by finite energy sum rules from the known sign of the photoproduction amplitude in the (3, 3) resonance region; the analysis has been carried out by Gilman,[9] who finds S positive. A third method consists of comparing Eq. (13) with an approximate expression for the $\pi^0 \to 2\gamma$ amplitude de-

rived by Goldberger and Treiman[10] (as corrected by Pagels[10]). These authors applied a pole dominance argument to proton Compton scattering dispersion relations, obtaining the relation

$$F \approx -4\pi\alpha \frac{\kappa_p}{g_r} \frac{1}{m_N} , \tag{14}$$

κ_p = proton anomalous magnetic moment = 1.79,

which gives a $\pi^0 \to 2\gamma$ rate of 2.0 eV, in fair agreement with experiment. Comparison of Eq. (14) with Eq. (13) again gives S positive. A fourth method which has been proposed[8] is to use Compton scattering data on protons to try to measure the interference of the pion exchange piece (proportional to F) with the nucleon and nucleon isobar exchange pieces. The problem with this proposal[11] is that one does not know whether to take the pion exchange piece in its Born approximation form, $tF/(t-\mu^2)$, or in the polology form, $\mu^2 F/(t-\mu^2)$. Since t is negative in the physical region, this uncertainty leads to a sign ambiguity and renders the method dubious. In any case, with fair certainty one learns from the first three methods that S is positive.

Armed with the experimental knowledge that $S = +0.5$, we can now use Eq. (12) to test various models of the hadrons which have been proposed. One very popular model is the triplet model, consisting of an SU_3-triplet of Fermions (p, n, λ) interacting by meson exchange. The charges of (p, n, λ) are $(Q, Q-1, Q-1)$ and the corresponding axial-vector couplings are $(g_p, g_n, g_\lambda) = (\frac{1}{2}, -\frac{1}{2}, 0)$. One immediately finds $S = \frac{1}{2} Q^2 - \frac{1}{2}(Q-1)^2 = Q - \frac{1}{2}$, and so $S = \frac{1}{2}$ requires $Q = 1$ [i. e., integral triplet charges (1, 0, 0)]. Note that the fractionally charged quark model has $Q = 2/3$, $S = 1/6$, and so the quark hypothesis is strongly excluded. Another integrally charged triplet model which is allowed is the Han-Nambu-Tavkhelidze[12] model, which has three triplets, S, U, B, with respective charges (1, 0, 0), (1, 0, 0), (0, -1, -1) and with axial-vector couplings $(\frac{1}{2}, -\frac{1}{2}, 0)$ for each triplet.

5. The ideas which we have developed can also be applied to the $\eta \to 2\gamma$ and the $X^0 \to 2\gamma$ decays.[13] Unfortunately, the experimental situation here is worse, and the theoretical situation is also worse, because the soft η and soft X^0 approximations involve a much larger extrapolation from the physical region than does the soft pion approximation. Nonetheless, pursuing this track, Glashow et al.[13] find a connection between the $\pi^0 \to 2\gamma$, $\eta \to 2\gamma$ and $X^0 \to 2\gamma$ decay rates, which predicts

$$\tau^{-1}(X^0 \to 2\gamma) \approx 350 \text{ keV} \quad \text{quark model } (Q = \tfrac{2}{3}),$$
$$\approx 120 \text{ keV} \quad \text{integrally charged } (Q = \pm 1) \text{ triplet model} . \tag{15}$$

Present experiments do not distinguish between the two alternatives in Eq. (15).

REFERENCES

1. J. Steinberger, Phys. Rev. 76, 1180 (1949).
2. D. G. Sutherland, Nuclear Phys. B2, 433 (1967).
3. S. L. Adler, Phys. Rev. 177, 2426 (1969); J. S. Bell and R. Jackiw, Nuovo Cimento 60, 47 (1969).
4. J. Schwinger, Phys. Rev. 82, 664 (1951); C. R. Hagen, Phys. Rev. 177, 2622 (1969); B. Zumino, in Proceedings of the Topical Conference on Weak Interactions, CERN, Geneva (1969), p. 361.
5. S. L. Adler, Ref. 3; S. L. Adler and W. A. Bardeen, Phys. Rev. (in press).
6. R. F. Dashen, private communication.
7. S. Devons et al., Phys. Rev. 184, 1356 (1969).
8. S. Okubo, Phys. Rev. 179, 1629 (1969).
9. F. J. Gilman, Phys. Rev. (in press).
10. M. L. Goldberger and S. B. Treiman, Nuovo Cimento 9, 451 (1958); H. Pagels, Phys. Rev. 158, 1566 (1967).
11. A. Hearn, private communication.
12. M. Y. Han and Y. Nambu, Phys. Rev. 139, B1006 (1965); A. Tavkhelidze, in High Energy Physics and Elementary Particles, International Atomic Energy Agency, Vienna (1965), p. 763.
13. S. Okubo (to be published); S.L. Glashow, R. Jackiw and S.S. Shei (to be published).

DISCUSSION

A. Dar: Recent measurements, by the Primakoff effect, of the π° decay rate give $\tau^{-1} = 11.2$ eV (± 10%). S.L.A.: Using $g^2/4\pi = 14.6$, the theoretical estimate is in the range 7.4 - 9.1 eV, but as indicated there is ∿ 20% uncertainty in the PCAC extrapolation. Incidentally, the new width is in still poorer agreement with the fractionally-charged quark model. V. Telegdi: Would one expect to see, in η-decay, evidence for a non E.M. isospin symmetry breaking interaction? S.L.A.: One can invent such an interaction to explain the 3π decay of the η. The extra terms I discussed will not affect this decay mode.

PHYSICAL REVIEW VOLUME 184, NUMBER 5 25 AUGUST 1969

Anomalous Commutators and the Triangle Diagram

STEPHEN L. ADLER
Institute for Advanced Study, Princeton, New Jersey 08540

AND

DAVID G. BOULWARE
University of Washington, Seattle, Washington 98105

(Received 19 March 1969)

We consider matrix elements of the axial-vector current in spinor electrodynamics, and develop the change in the usual reduction formalism caused by the presence of the axial-vector-current–two-photon triangle diagram. When at most one photon is reduced in from the external states, we are able to characterize the anomalous behavior of the triangle diagram entirely in terms of a consistent set of anomalous field-current and current-current commutators.

It has recently been shown[1] that the axial-vector current in spinor electrodynamics does not satisfy the usual divergence equation

$$\partial^\mu j_\mu{}^5(x)=2im_0 j^5(x),$$

where

$$j_\mu{}^5(x)=\bar\psi(x)\gamma_\mu\gamma_5\psi(x), \quad j^5(x)=\bar\psi(x)\gamma_5\psi(x), \tag{1}$$

expected from naive use of the equations of motion. Rather, because of the presence of the triangle diagram shown in Fig. 1, the axial-vector current satisfies the anomalous divergence condition

$$\partial^\mu j_\mu{}^5(x)=2im_0 j^5(x)+(\alpha_0/4\pi)F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho}, \tag{2}$$

with $F^{\xi\sigma}$ the unrenormalized electromagnetic field-strength tensor. Because radiative corrections to the basic triangle diagram (Fig. 2) involve axial-vector loops with at least five vertices, and because these larger loops satisfy the usual axial-vector Ward identity, Eq. (2) is an *exact* equation, valid to all orders in perturbation theory.[2]

In the present paper we explore further consequences of the singular behavior of the triangle diagram in spinor electrodynamics. Although the anomalous divergence phenomenon appears in all matrix elements of the axial-vector current, we will consider explicitly only the axial-vector-current–two-photon matrix element $\langle 0|j_\mu{}^5|k_1,\epsilon_1;k_2,\epsilon_2\rangle$, which is described in lowest order by the graph of Fig. 1. (Here k_1, k_2 and ϵ_1, ϵ_2 denote, respectively, the four-momenta and polarizations of the two photons.) First, we develop the reduction formalism for the triangle graph. When one photon is reduced in, we are able to characterize the anomalous behavior of the triangle graph entirely in terms of anomalous commutators of the electromagnetic field with the axial-vector current ("seagulls") and of the electromagnetic current with the axial-vector current ("Schwinger terms"). We check that the various commutators which we obtain are consistent with each other, with the equations of motion, and with the electromagnetic-field canonical commutation relations. These formal considerations indicate that the equations obtained from explicit study of the matrix element $\langle 0|j_\mu{}^5|k_1,\epsilon_1;k_2,\epsilon_2\rangle$ can be applied unchanged to the matrix element $\langle A|j_\mu{}^5|B\rangle$, with A and B arbitrary, when at most one photon is reduced in from the external states. Using the anomalous commutation relations, we complete the heuristic verification that the quantity $\tilde{Q}^5$ introduced in I is the chiral generator in massless electrodynamics. Finally, we show that when both photons are pulled in, one cannot represent the triangle graph by a reduction formula containing a time-ordered product with the usual properties.

To study the reduction formula for the triangle graph with one photon pulled in, we use the equation[3]

$$\begin{aligned}\langle 0|j_\mu{}^5(0)|k_1,\epsilon_1;k_2,\epsilon_2\rangle&[(2\pi)^3 2k_{10}(2\pi)^3 2k_{20}]^{1/2}\\ &=-i\epsilon_1{}^\sigma\int d^4x\, e^{-ik_1\cdot x}\\ &\qquad\times\Box_x\langle 0|T(j_\mu{}^5(0)A_\sigma(x))|k_2,\epsilon_2\rangle[(2\pi)^3 2k_{20}]^{1/2}\\ &=-i\epsilon_1{}^\sigma\epsilon_2{}^\rho[e_0{}^2/(2\pi)^4]R_{\sigma\rho\mu}(k_1,k_2),\end{aligned} \tag{3}$$

where A_σ is the photon field and $R_{\sigma\rho\mu}(k_1,k_2)$ is the explicit expression for the lowest-order triangle graph given in Eqs. (17) and (18) of I. Bringing $\Box_x$ inside the time-ordered product (using the usual rules[4] for differentiating time-ordered products), we find

$$\int d^4x\, e^{-ik_1\cdot x}\Box_x\langle 0|T(j_\mu{}^5(0)A_\sigma(x))|k_2,\epsilon_2\rangle =A_{\mu\sigma}k_{10}+B_{\mu\sigma}+C_{\mu\sigma}(k_{10}), \tag{4}$$

[1] S. L. Adler, Phys. Rev. 177, 2426 (1969). This paper will hereafter be referred to as I. See also J. Schwinger, *ibid.* 82, 664 (1951), Sec. V; C. R. Hagen, *ibid.* 177, 2622 (1969); R. Jackiw and K. Johnson, *ibid.* 182, 1457 (1969); B. Zumino (unpublished). As in I, we use the notation and metric conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Co., New York, 1965), pp. 377-390. In particular, we use $\epsilon_{0123}=\epsilon^{123}=1$.

[2] S. L. Adler and W. A. Bardeen, Phys. Rev. 182, 1515 (1969). Note that the anomalous divergence term can be rewritten in terms of finite quantities as $(\alpha/4\pi)F_r{}^{\xi\sigma}F_r{}^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$, where $F_r{}^{\xi\sigma}$ is the renormalized electromagnetic field-strength tensor.

[3] Since in Eqs. (3)-(7) we work to lowest order only, we omit the wave-function renormalization factor from Eq. (3).

[4] S. L. Adler and R. F. Dashen, *Current Algebras* (W. A. Benjamin, Inc., New York, 1968), Eq. (2.7).

with[5]

$$
\begin{aligned}
A_{\mu\sigma}&=i\int d^4x\, e^{i\mathbf{k}_1\cdot\mathbf{x}}\delta(x_0)\langle 0|[\dot{A}_\sigma(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle,\\
B_{\mu\sigma}&=\int d^4x\, e^{i\mathbf{k}_1\cdot\mathbf{x}}\delta(x_0)\langle 0|[A_\sigma(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle,\\
C_{\mu\sigma}(k_{10})&=e_0\int d^4x\, e^{-ik_1\cdot x}\langle 0|T(j_\mu{}^5(0)j_\sigma(x))|k_2,\epsilon_2\rangle,\\
\dot{A}_\sigma(x)&\equiv\frac{\partial}{\partial x_0}A_\sigma(x),\quad j_\sigma(x)\equiv\bar{\psi}(x)\gamma_\sigma\psi(x).
\end{aligned}
\tag{5}
$$

Provided that the time-ordered product in $C_{\mu\nu}$ is not too singular, in the limit as $k_{10}\to\infty$, the function $C_{\mu\nu}(k_{10})$ has the Bjorken[6]-Johnson-Low[7] behavior

$$
\begin{aligned}
C_{\mu\nu}(k_{10})=\frac{-ie_0}{k_{10}}\int d^4x\, e^{i\mathbf{k}_1\cdot\mathbf{x}}\delta(x_0)&\\
\times\langle 0|[j_\sigma(\mathbf{x}),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle&+O[(\ln k_{10})^\beta/k_{10}{}^2],
\end{aligned}
\tag{6}
$$

indicating that the equal-time commutators $[A_\sigma(x),j_\mu{}^5(0)]$, $[\dot{A}_\sigma(x),j_\mu{}^5(0)]$, and $[j_\sigma(x),j_\mu{}^5(0)]$ are to be identified, respectively, with the parts of $R_{\sigma\rho\mu}$ behaving like k_{10}, 1, and $k_{10}{}^{-1}$ as k_{10} becomes infinite. From Eqs. (17) and (18) of I, we find

$$
\begin{aligned}
&\epsilon_2{}^\rho R_{\sigma\rho\mu}(k_1,k_2)\\
&\quad=4\pi^2(k_2{}^\tau\epsilon_2{}^\rho-k_2{}^\rho\epsilon_2{}^\tau)\{\epsilon_{\tau\sigma\rho\mu}+g_{\sigma0}\epsilon_{0\tau\rho\mu}-g_{\rho0}\epsilon_{0\tau\sigma\mu}\\
&\qquad+k_{10}{}^{-1}[\tfrac{1}{2}(1-g_{\sigma0})(k_{2\sigma}\epsilon_{0\tau\rho\mu}+k_{20}\epsilon_{\sigma\tau\rho\mu})\\
&\qquad+g_{\sigma0}(1-g_{\eta0})k_1{}^\eta\epsilon_{\eta\tau\rho\mu}-g_{\rho0}(1-g_{\eta0})k_1{}^\eta\epsilon_{\eta\tau\sigma\mu}\\
&\qquad+(\text{terms which vanish when }\sigma=0\text{ or }\mu=0)]\}\\
&\qquad\qquad+O[(\ln k_{10})^\beta/k_{10}{}^2].
\end{aligned}
\tag{7}
$$

Comparing Eq. (7) with Eqs. (5) and (6), we find the equal-time commutation relations[8]

$$
\begin{aligned}
&[A_\sigma(x),j_\mu{}^5(y)]=[\dot{A}_0(x),j_\mu{}^5(y)]=0,\\
&[\dot{A}_r(x),j_0{}^5(y)]=(-2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})B^r(y),\\
&[\dot{A}_r(x),j_s{}^5(y)]=(i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})\epsilon^{rst}E^t(y),\\
&[j_0(x),j_0{}^5(y)]=(-ie_0/2\pi^2)\mathbf{B}(y)\cdot\nabla_{\mathbf{x}}\delta^3(\mathbf{x}-\mathbf{y}),\\
&[j_r(x),j_0{}^5(y)]=(-ie_0/4\pi^2)[\mathbf{E}(x)\times\nabla_{\mathbf{y}}\delta^3(\mathbf{x}-\mathbf{y})]^r,\\
&[j_0(x),j_s{}^5(y)]=(ie_0/4\pi^2)[\mathbf{E}(y)\times\nabla_{\mathbf{x}}\delta^3(\mathbf{x}-\mathbf{y})]^s,
\end{aligned}
\tag{8}
$$

[5] We have suppressed the dependence of $A_{\mu\sigma},\ldots,C_{\mu\sigma}$ on $\mathbf{k}_1$ and k_2.

[6] J. D. Bjorken, Phys. Rev. 148, 1467 (1968).

[7] K. Johnson and F. E. Low, Progr. Theoret. Phys. (Kyoto) Suppl. 37-38, 74 (1966).

[8] We remind the reader that since we have deduced the commutators of Eq. (8) from the triangle graph alone, without considering other graphs, we have not yet ruled out the presence of additional terms in the field-current or current-current commutators of higher order than α_0 or e_0, respectively. However, the consistency argument of Eqs. (23)–(30) below suggests that such terms, if they occur at all, are at worst Schwinger terms and seagulls of the usual type, which cancel against each other when vector or axial-vector divergences are taken.

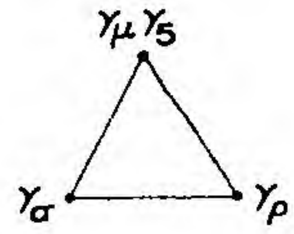


FIG. 1. Axial-vector triangle diagram which leads to the extra term in Eq. (2).

with

$$
\begin{aligned}
B^t(x)&=[\nabla\times\mathbf{A}(x)]^t=\epsilon^{rst}\frac{\partial}{\partial x^r}A^s(x),\\
E^t(x)&=-\dot{A}^t(x)-\frac{\partial}{\partial x^t}A^0(x),\\
\epsilon^{123}&=1.
\end{aligned}
\tag{9}
$$

We have only listed the current-current commutators containing at least one time component, since these are the only ones which appear when divergences with respect to the vector or axial-vector indices (σ or μ) are brought inside the time-ordered product in Eq. (5). All of the nonvanishing commutators in Eq. (8) are anomalous in the sense that if they are calculated by naive use of canonical commutation relations they vanish.

It is easy to check that the anomalous commutation relations of Eq. (8), together with the reduction formula of Eqs. (4) and (5), correctly reproduce the known divergence properties of the lowest-order triangle diagram. Consistent with our assumption that the time-ordered product $C_{\mu\nu}$ is not too singular, and obeys the Bjorken-Johnson-Low asymptotic formula, we use the usual formulas[4] for differentiation of the time-ordered product,

$$
\begin{aligned}
\partial_y{}^\mu T(j_\mu{}^5(y)j_\sigma(x))&=T(\partial_y{}^\mu j_\mu{}^5(y)j_\sigma(x))\\
&\quad+\delta(y^0-x^0)[j_0{}^5(y),j_\sigma(x)],\\
\partial_x{}^\sigma T(j_\mu{}^5(y)j_\sigma(x))&=T(j_\mu{}^5(y)\partial_x{}^\sigma j_\sigma(x))\\
&\quad+\delta(x^0-y^0)[j_0(x),j_\mu{}^5(y)].
\end{aligned}
\tag{10}
$$

To check gauge invariance for the photon which has been reduced in, we multiply Eq. (4) by $k_1{}^\sigma$. Using vector-current conservation ($\partial^\sigma j_\sigma=0$) and Eq. (10) to evaluate $k_1{}^\sigma C_{\mu\sigma}(k_{10})$, we find

$$
\begin{aligned}
&k_1{}^\sigma\int d^4x\, e^{-ik_1\cdot x}\Box_x\langle 0|T(j_\mu{}^5(0)A_\sigma(x))|k_2,\epsilon_2\rangle\\
&\quad=k_1{}^\sigma\int d^4x\, e^{i\mathbf{k}_1\cdot\mathbf{x}}\delta(x^0)\langle 0|[\dot{A}_\sigma(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle\\
&\quad\quad-ie_0\int d^4x\, e^{i\mathbf{k}_1\cdot\mathbf{x}}\delta(x^0)\langle 0|[j_0(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle.
\end{aligned}
\tag{11}
$$

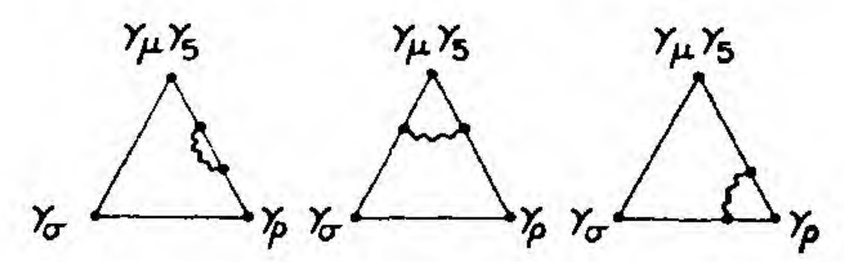


FIG. 2. Typical second-order radiative corrections to the triangle diagram.

Using the commutators of Eq. (8), one can easily see that the right-hand side of Eq. (11) vanishes. To check the axial-vector divergence of the triangle, we multiply Eq. (4) by $-(k_1+k_2)^\mu$. Using the axial-vector-current divergence equation (2) and Eq. (10) to evaluate $(k_1+k_2)^\mu C_{\mu\sigma}(k_{10})$, we find

$$
\begin{aligned}
-(k_1+k_2)^\mu \int d^4x\, e^{-ik_1\cdot x}\Box_x\langle 0|T(j_\mu{}^5(0)A_\sigma(x))|k_2,\epsilon_2\rangle \\
= -ie_0\int d^4x\, e^{-ik_1\cdot x}\langle 0|T([2im_0j^5(0) \\
+(\alpha_0/4\pi)F^{\xi\sigma}(0)F^{\tau\eta}(0)\epsilon_{\xi\sigma\tau\eta}]j_\sigma(x))|k_2,\epsilon_2\rangle \\
+\int d^4x\, e^{ik_1\cdot x}\delta(x_0) \\
\times\{-(k_1+k_2)^\mu\langle 0|[A_\sigma(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle \\
+ie_0\langle 0|[j_\sigma(x),j_0{}^5(0)]|k_2,\epsilon_2\rangle\}. \quad (12a)
\end{aligned}
$$

Since we are only working to lowest order (order $e_0{}^2$), the anomalous divergence term proportional to $e_0\alpha_0F^{\xi\sigma}F^{\tau\eta}\epsilon_{\xi\sigma\tau\eta}$ makes no contribution. However, the anomalous commutator terms in curly brackets may be evaluated from Eq. (8), and they give

$$
\begin{aligned}
\int d^4x\, e^{ik_1\cdot x}\delta(x_0)\{-(k_1+k_2)^\mu\langle 0|[A_\sigma(x),j_\mu{}^5(0)]|k_2,\epsilon_2\rangle \\
+ie_0\langle 0|[j_\sigma(x),j_0{}^5(0)]|k_2,\epsilon_2\rangle\}[(2\pi)^3 2k_{20}]^{1/2} \\
=-\epsilon_2{}^\rho(e_0{}^2/2\pi^2)k_1{}^\xi k_2{}^\tau\epsilon_{\xi\sigma\tau\rho}. \quad (12b)
\end{aligned}
$$

When multiplied by $\epsilon_1{}^\sigma$, Eq. (12b) is identical with the matrix element $\langle 0|(\alpha_0/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}|k_1,\epsilon_1;k_2,\epsilon_2\rangle \times[(2\pi)^3 2k_{10}(2\pi)^3 2k_{20}]^{1/2}$ which comes from the anomalous term in Eq. (2) if we calculate the divergence of $\langle 0|j_\mu{}^5|k_1,\epsilon_1;k_2,\epsilon_2\rangle$ directly, *before* reducing in one photon. We see then that the reduction formula of Eqs. (4) and (5), combined with the anomalous commutators of Eq. (8), correctly characterizes the anomalous axial-vector index divergence of the triangle diagram. As Jackiw and Johnson[1] have particularly emphasized, in the reduction formula the anomalous divergence term $k_1{}^\xi k_2{}^\tau\epsilon_{\xi\sigma\tau\rho}$ arises from the failure of the "Schwinger term" $[j_\sigma,j_0{}^5]$ and the "seagull" $[\dot{A}_\sigma,j_\mu{}^5]$ to cancel. {As a point of consistency, we note that the pseudoscalar–two-photon triangle $R_{\sigma\rho}$ [defined in Eq. (19) of I] has the asymptotic behavior $R_{\sigma\rho}(k_1,k_2)\to 0$ as $k_{10}\to\infty$. Thus the usual equal-time commutation relations

$$[A_\sigma(x),j^5(y)]=[\dot{A}_\sigma(x),j^5(y)]=0 \quad (13)$$

remain valid, and no extra seagull terms are picked up when the one-photon reduction formula is applied to the matrix element $\langle 0|2im_0j^5|k_1,\epsilon_1;k_2,\epsilon_2\rangle$.}

We proceed next to check whether the commutation relations of Eqs. (8) and (13) are formally consistent with each other, with the equations of motion, and with the usual electromagnetic-field canonical commutation relations. In the Feynman gauge, the electromagnetic-field equations of motion and commutation relations are

$$
\begin{aligned}
&\Box A_\mu=\ddot{A}_\mu-\nabla^2A_\mu=e_0j_\mu, \\
&[A^\lambda(x),A^\sigma(y)]|_{x^0=y^0}=[\dot{A}^\lambda(x),\dot{A}^\sigma(y)]|_{x^0=y^0}=0, \quad (14) \\
&[A^\lambda(x),\dot{A}^\sigma(y)]|_{x^0=y^0}=-ig^{\lambda\sigma}\delta^3(\mathbf{x}-\mathbf{y}).
\end{aligned}
$$

We also need the divergance equations satisfied by the currents $j_\mu(\mathbf{x},t)$ and $j_\mu{}^5(\mathbf{x},t)$,

$$
\begin{aligned}
&\frac{\partial}{\partial t}j_0+\nabla\cdot\mathbf{j}=0, \\
&\frac{\partial}{\partial t}j_0{}^5+\nabla\cdot\mathbf{j}^5=2im_0j^5+(2\alpha_0/\pi)\mathbf{E}\cdot\mathbf{B},
\end{aligned} \quad (15)
$$

with **E** and **B** given, of course, by Eq. (9). We proceed to combine Eqs. (14) and (15) with Eqs. (8) and (13). All the commutators which we write down are at equal time, with $x^0=y^0=t$.

(i) From $[A_\sigma(x),j_0{}^5(y)]=0$, we deduce[9]

$$[\dot{A}_\sigma(x),j_0{}^5(y)]+[A_\sigma(x),(\partial/\partial t)j_0{}^5(y)]=0. \quad (16)$$

On substituting Eq. (15) for $(\partial/\partial t)j_0{}^5(y)$ and using $[A_\sigma(x),j^5(y)]=[A_\sigma(x),\mathbf{j}^5(y)]=0$, we find

$$[\dot{A}_\sigma(x),j_0{}^5(y)]=-[A_\sigma(x),(2\alpha_0/\pi)\mathbf{E}(y)\cdot\mathbf{B}(y)]. \quad (17)$$

Using the canonical commutation relations of Eq. (14), we then get

$$[\dot{A}_0(x),j_0{}^5(y)]=0, \quad (18)$$

$$[\dot{A}_r(x),j_0{}^5(y)]=(-2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})B^r(y), \quad (19)$$

in agreement with Eq. (8).

(ii) From $[\dot{A}_0(x),j_0{}^5(y)]=0$, we deduce

$$[\ddot{A}_0(x),j_0{}^5(y)]+[\dot{A}_0(x),(\partial/\partial t)j_0{}^5(y)]=0. \quad (20)$$

Substituting Eq. (15) for $(\partial/\partial t)j_0{}^5(y)$ and Eq. (14) for $\ddot{A}_0(x)$, and using the commutators $[A_0(x),j_0{}^5(y)] =[\dot{A}_0(x),j^5(y)]=[\dot{A}_0(x),\mathbf{j}^5(y)]=0$, we find

$$
\begin{aligned}
[e_0j_0(x),j_0{}^5(y)]&=-[\dot{A}_0(x),(2\alpha_0/\pi)\mathbf{E}(y)\cdot\mathbf{B}(y)] \\
&=(-2i\alpha_0/\pi)\mathbf{B}(y)\cdot\nabla_x\delta^3(\mathbf{x}-\mathbf{y}), \quad (21)
\end{aligned}
$$

that is,

$$[j_0(x),j_0{}^5(y)]=(-ie_0/2\pi^2)\mathbf{B}(y)\cdot\nabla_x\delta^3(\mathbf{x}-\mathbf{y}), \quad (22)$$

in accord with Eq. (8).

(iii) From $[\dot{A}_r(x),j_0{}^5(y)]=-(2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})B^r(y)$, we find

$$
\begin{aligned}
&[\ddot{A}_r(x),j_0{}^5(y)]+[\dot{A}_r(x),(\partial/\partial t)j_0{}^5(y)] \\
&\quad=(-2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})\dot{B}^r(y) \\
&\quad=(2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})[\nabla_y\times\mathbf{E}(y)]^r. \quad (23)
\end{aligned}
$$

[9] We use here the method of D. G. Boulware and L. S. Brown, Phys. Rev. 156, 1724 (1967).

Substituting for $\ddot{A}_r(x)$ and $(\partial/\partial t)j_0{}^5(y)$ as before, we find

$$\begin{aligned}[e_0 j_r(x),j_0{}^5(y)]-[\dot{A}_r(x),\nabla_y\cdot\mathbf{j}^5(y)]\\ =(2i\alpha_0/\pi)\delta^3(\mathbf{x}-\mathbf{y})[\nabla_y\times\mathbf{E}(y)]^r\\ -[\dot{A}_r(x),(2\alpha_0/\pi)\mathbf{E}(y)\cdot\mathbf{B}(y)]\\ =(-2i\alpha_0/\pi)[\mathbf{E}(x)\times\nabla_y\delta^3(\mathbf{x}-\mathbf{y})]^r.\end{aligned}\tag{24}$$

Using Eq. (8) to evaluate $[e_0 j_r(x),j_0{}^5(y)]$ and $-[\dot{A}_r(x),\nabla\cdot\mathbf{j}^5(y)]$, we see that Eq. (24) is satisfied.

(iv) From $[j_0(x),j_0{}^5(y)]=-(ie_0/2\pi^2)\mathbf{B}(y)\cdot\nabla_x\delta^3(\mathbf{x}-\mathbf{y})$, we find

$$\begin{aligned}[(\partial/\partial t)j_0(x),j_0{}^5(y)]+[j_0(x),(\partial/\partial t)j_0{}^5(y)]\\ =(-ie_0/2\pi^2)\dot{\mathbf{B}}(y)\cdot\nabla_x\delta^3(\mathbf{x}-\mathbf{y})\\ =(ie_0/2\pi^2)[\nabla_y\times\mathbf{E}(y)]\cdot\nabla_x\delta^3(\mathbf{x}-\mathbf{y}).\end{aligned}\tag{25}$$

Substituting Eq. (15) for $(\partial/\partial t)j_0(x)$ and $(\partial/\partial t)j_0{}^5(x)$ gives[10]

$$\begin{aligned}-[\nabla_x\cdot\mathbf{j}(x),j_0{}^5(y)]-[j_0(x),\nabla_y\cdot\mathbf{j}^5(y)]\\ =(ie_0/2\pi^2)[\nabla_y\times\mathbf{E}(y)]\cdot\nabla_x\delta^3(x-y).\end{aligned}\tag{26}$$

Using Eq. (8) to calculate the commutators on the left-hand side, we find that Eq. (26) is satisfied.

(v) Finally, to check the consistency of quantization in the Feynman gauge, we must verify that

$$L\equiv\dot{A}_0+\nabla\cdot\mathbf{A}\tag{27}$$

and $\dot{L}$ remain dynamically independent of the axial-vector current. That is, we must verify that

$$[L(x),j_\mu{}^5(y)]=0\tag{28}$$

and that

$$[\dot{L}(x),j_\mu{}^5(y)]=0.\tag{29}$$

Equation (28) follows immediately from the first line of Eq. (8). To check Eq. (29), we substitute Eq. (14) for $\ddot{A}_0$ and use $[A_0(x),j_\mu{}^5(y)]=0$, giving

$$\begin{aligned}[\dot{L}(x),j_\mu{}^5(y)]=[e_0 j_0(x),j_\mu{}^5(y)]\\ +[\nabla_x\cdot\dot{\mathbf{A}}(x),j_\mu{}^5(y)].\end{aligned}\tag{30}$$

Substituting commutators from Eq. (8) then shows that the right-hand side of Eq. (30) vanishes.

We conclude that the commutation relations of Eq. (8), which were obtained from the triangle graph in lowest-order perturbation theory, are consistent with the equations of motion and canonical commutation relations of Eqs. (14) and (15). Moreover, the fact that Eq. (19) for $[\dot{A}_r(x),j_0{}^5(y)]$ and Eq. (22) for $[j_0(x),j_0{}^5(y)]$ were *deduced* from simpler, exact commutators[11] and equations of motion[2] suggests that Eqs. (19) and (22) are themselves exact to all orders of perturbation theory.[12] The values given in Eq. (8) for $[\dot{A}_r(x),j_s{}^5(y)]$, $[j_r(x),j_0{}^5(y)]$, and $[j_0(x),j_s{}^5(y)]$ cannot, on the other hand, be deduced from the consistency argument of Eqs. (23)–(30). To see this, we note that Eqs. (24), (26), and (30) [as well as the reduction formulas (11) and (12)] are all unchanged if we modify these commutators to read

$$\begin{aligned}[\dot{A}_r(x),j_s{}^5(y)]&=\frac{i\alpha_0}{\pi}\delta^3(\mathbf{x}-\mathbf{y})\epsilon^{rst}E^t(y)\\ &\quad-ie_0\delta^3(\mathbf{x}-\mathbf{y})S^{rs}(y),\\ [j_r(x),j_0{}^5(y)]&=\frac{-ie_0}{4\pi^2}[\mathbf{E}(x)\times\nabla_y\delta^3(\mathbf{x}-\mathbf{y})]^r\\ &\quad+i\frac{\partial}{\partial y^s}[\delta^3(\mathbf{x}-\mathbf{y})S^{rs}(y)],\\ [j_0(x),j_s{}^5(y)]&=\frac{ie_0}{4\pi^2}[\mathbf{E}(y)\times\nabla_x\delta^3(\mathbf{x}-\mathbf{y})]^s\\ &\quad-i\frac{\partial}{\partial x^r}[\delta^3(\mathbf{x}-\mathbf{y})S^{rs}(y)],\end{aligned}\tag{31}$$

with $S^{rs}(y)$ a pseudotensor operator. In other words, the consistency check of Eqs. (23)–(30) does not rule out the possibility that higher orders of perturbation theory may modify Eq. (8) by adding Schwinger terms and seagulls of the usual type,[13] which cancel against each other [as in Eqs. (11) and (12)] when vector or axial-vector divergences are taken. It is expected[13] on general grounds that the commutator $[\dot{A}_r(x),j_s{}^5(y)]$ does not involve derivatives of the δ function and that the commutators $[j_r(x),j_0{}^5(y)]$ and $[j_0(x),j_s{}^5(y)]$ do

[10] We have used $[j_0(x),\mathbf{E}(y)\cdot\mathbf{B}(y)]=0$, which follows from $[j_0(x),A_\sigma(y)]=[j_0(x),\dot{A}_\sigma(y)]=0$. Note that $[j_0(x),\ddot{A}_\sigma(y)]=0$ can be derived from $[j_0(x),A_\sigma(y)]=0$ and the divergence equation $\partial j_0(x)/\partial t+\nabla_x\cdot\mathbf{j}=0$, in the same way that we derived Eq. (19).

[11] The commutation relations $[A_\sigma(x),j_\mu{}^5(y)]=[A_\sigma(x),j^5(y)]=[\dot{A}_0(x),j_\mu{}^5(y)]=[\dot{A}_0(x),j^5(y)]=0$ can be proved to all orders in perturbation theory by the Bjorken-Johnson-Low method, using the Weinberg asymptotic rules discussed in Chap. 19 of Bjorken and Drell (Ref. 1). Let $T_{\sigma\mu 2f b}(k_1,\ldots)$ be an arbitrary amplitude involving an external photon of polarization σ and four-momentum k_1, an axial-vector current $j_\mu{}^5$ with four-momentum $-k_1+\Delta$, $2f$ external fermions, and b additional external photons. Because of charge-conjugation invariance, we cannot have $2f=b=0$. When $f>0$ or $b>1$, the *asymptotic coefficient* α associated with T, as $k_{10}\to\infty$, can never be greater than zero. When $f=0$ and $b=1$, the superficial asymptotic coefficient is 1 (the graph is linearly divergent), but gauge invariance implies that the photon b must couple through its field-strength tensor, and this reduces the effective α to zero. Thus α for T can never be greater than zero, and since T is arbitrary, this statement holds for all subgraphs of T as well. We conclude that $T_{\sigma\mu 2f b}(k_1,\ldots)\sim(\ln k_{10})^\beta$ as $k_{10}\to\infty$, and since $k_1{}^\sigma T_{\sigma\mu 2f b}(k_1,\ldots)=0$ by gauge invariance, this means that $T_{0\mu 2f b}(k_1,\ldots)\sim k_{10}{}^{-1}(\ln k_{10})^\beta$. Comparing with Eq. (5), we conclude that $[A_\sigma(x),j_\mu{}^5(y)]=[\dot{A}_0(x),j_\mu{}^5(y)]=0$. An identical argument holds with $j_\mu{}^5$ replaced by j^5.

[12] We believe that Eqs. (19) and (22) are exact when sandwiched between normalizable states $\langle a|$ and $|b\rangle$. We make no claims about matrix elements involving non-normalizable states such as $j_\sigma(x)|a\rangle$ or $j_\mu{}^5(y)|a\rangle$ and, in particular, we do *not* demand that the commutators of Eq. (8) satisfy the Jacobi identity. (They do *not*.) For a discussion of Jacobi-identity breakdown, see Johnson and Low (Ref. 7).

[13] See Adler and Dashen (Ref. 4), Chap. 3; Boulware and Brown (Ref. 9); D. G. Boulware, Phys. Rev. 172, 1625 (1968).

not involve derivatives of the δ function higher than the first. Under this assumption, Eq. (31) represents the most general form for these commutators consistent with Eqs. (14) and (15).

Using Eq. (19), we can easily complete the argument sketched in I that the operator

$$\bar{Q}^5=\int d^3x\left[j_0{}^5(x)+(\alpha_0/\pi)\mathbf{A}(x)\cdot\boldsymbol{\nabla}_x\times\mathbf{A}(x)\right] \quad (32)$$

is the conserved generator of γ_5 transformations in massless electrodynamics. In I it was shown that

$$\frac{d}{dt}\bar{Q}^5=0,\quad [\bar{Q}^5,\psi(y)]=-i\gamma_5\psi(y). \quad (33)$$

We now show that $\bar{Q}^5$ commutes with the photon field variables. From the first line of Eq. (8) we find

$$[\bar{Q}^5,A_\sigma(y)]=[\bar{Q}^5,\dot{A}_0(y)]=0, \quad (34a)$$

while from Eq. (19) we find[14]

$$[\bar{Q}^5,\dot{A}_r(y)]=\left[\int d^3x\, j_0{}^5(x),\dot{A}_r(y)\right]+\left[\int d^3x\,(\alpha_0/\pi)\mathbf{A}(x)\cdot\boldsymbol{\nabla}_x\times\mathbf{A}(x),\dot{A}_r(y)\right]$$
$$=\frac{2i\alpha_0}{\pi}B^r(y)-\frac{2i\alpha_0}{\pi}B^r(y)=0, \quad (34b)$$

as promised.

Finally, we will show that when two photons are pulled in, the triangle graph cannot be represented by a reduction formula containing a time-ordered product with the usual properties. When two photons are pulled in, Eq. (3) is replaced by[15]

$$\int d^4x d^4y\, e^{-ik_1\cdot x}e^{-ik_2\cdot y}\times\Box_x\Box_y\langle 0|T(j_\mu{}^5(0)A_\sigma(x)A_\rho(y))|0\rangle$$
$$=i[e_0{}^2/(2\pi)^4]R_{\sigma\rho\mu}(k_1,k_2). \quad (35)$$

Bringing $\Box_x$ and $\Box_y$ inside the time-ordered product on the left-hand side of Eq. (35) gives[16]

$$\int d^4x d^4y\, e^{-ik_1\cdot x}e^{-ik_2\cdot y}\times\Box_x\Box_y\langle 0|T(j_\mu{}^5(0)A_\sigma(x)A_\rho(y))|0\rangle$$
$$=C_{\mu\sigma\rho}(k_{10},k_{20})+S_{\mu\sigma\rho}(k_{10},k_{20}), \quad (36)$$
$$C_{\mu\sigma\rho}(k_{10},k_{20})=e_0{}^2\int d^4x d^4y\, e^{-ik_1\cdot x}e^{-ik_2\cdot y}\times\langle 0|T(j_\mu{}^5(0)j_\sigma(x)j_\rho(y))|0\rangle.$$

The "time-ordered product" $C_{\mu\sigma\rho}$ contains all of the dynamical singularities of the matrix element, but in addition there is a polynomial in k_1 and k_2, which we have labeled $S_{\mu\sigma\rho}$, arising from anomalous commutators of A and $\dot{A}$ with the currents. If the time-ordered product $C_{\mu\sigma\rho}$ were of the usual type, then it would have the Bjorken-Johnson-Low behavior in the limits as k_{10}, k_{20}, or $k_{10}-k_{20}$ become infinite. That is, we would have

$$C_{\mu\sigma\rho}(k_{10},k_{20})\xrightarrow[k_{10}\to\infty,\,k_{20}\text{ fixed}]{}\frac{-ie_0{}^2}{k_{10}}\int d^4x d^4y\times e^{ik_1\cdot x}\delta(x_0)e^{-ik_2\cdot y}\langle 0|T([j_\sigma(x),j_\mu{}^5(0)]j_\rho(y))|0\rangle+O((\ln k_{10})^\beta/k_{10}{}^2),$$

$$C_{\mu\sigma\rho}(k_{10},k_{20})\xrightarrow[k_{20}\to\infty,\,k_{10}\text{ fixed}]{}\frac{-ie_0{}^2}{k_{20}}\int d^4x d^4y\times e^{-ik_1\cdot x}e^{ik_2\cdot y}\delta(y_0)\langle 0|T([j_\rho(y),j_\mu{}^5(0)]j_\sigma(x))|0\rangle+O((\ln k_{20})^\beta/k_{20}{}^2), \quad (37)$$

$$C_{\mu\sigma\rho}(k_{10},k_{20})\xrightarrow[k_{10}-k_{20}\to\infty,\,k_{10}+k_{20}\text{ fixed}]{}\frac{-ie_0{}^2}{k_{10}-k_{20}}\int d^4x d^4y\times e^{-\frac{1}{2}i(k_1+k_2)\cdot(x+y)}e^{\frac{1}{2}i(\mathbf{k}_1-\mathbf{k}_2)\cdot(\mathbf{x}-\mathbf{y})}\delta(\tfrac{1}{2}(x_0-y_0))\times\langle 0|T([j_\sigma(x),j_\rho(y)]j_\mu{}^5(0))|0\rangle+O[(\ln(k_{10}-k_{20}))^\beta/(k_{10}-k_{20})^2].$$

According to Eqs. (36) and (37), all terms in $R_{\sigma\rho\mu}$ which either approach constants or diverge linearly in the three limits must be contained entirely in the polynomial S. In Eq. (7) we saw that as $k_{10}\to\infty$, with k_{20} fixed, $R_{\sigma\rho\mu}$ approaches a nonzero finite limit and, by Bose symmetry, the same statement holds for the limit $k_{20}\to\infty$, with k_{10} fixed. In I it was shown that in the limit $k_{10}-k_{20}\to\infty$, with $k_{10}+k_{20}$ fixed, $R_{\sigma\rho\mu}$ diverges linearly (i.e., behaves as finite coefficient times $k_{10}-k_{20}$). Clearly, these three limiting behaviors *cannot* be described by a polynomial in k_{10} and k_{20}, which means that $C_{\mu\sigma\rho}$ cannot vanish in all three of the limits in Eq. (37). Thus, the time-ordered product appearing in the two-photon reduction formula is not of the usual type.

We wish to thank L. S. Brown, R. Jackiw, and S. B. Treiman for helpful conversations.

[14] Equation (34) and Eqs. (16) and (19) may be combined into the simple observation that $[\bar{Q}^5,A_r]=0$ and $d\bar{Q}^5/dt=0$ implies $[\bar{Q}^5,\dot{A}_r]=0$.

[15] Again, we neglect the photon wave-function renormalization.

[16] We have suppressed the dependence of $C_{\mu\sigma\rho}$ and $S_{\mu\sigma\rho}$ on $\mathbf{k}_1$ and $\mathbf{k}_2$.

PHYSICAL REVIEW VOLUME 182, NUMBER 5 25 JUNE 1969

Absence of Higher-Order Corrections in the Anomalous Axial-Vector Divergence Equation

STEPHEN L. ADLER AND WILLIAM A. BARDEEN*
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 24 February 1969)

We consider two simple field-theoretic models, (a) spinor electrodynamics and (b) the σ model with the Polkinghorne axial-vector current, and show in each case that the axial-vector current satisfies a simple anomalous divergence equation exactly to all orders of perturbation theory. We check our general argument by an explicit calculation to second order in radiative corrections. The general argument is made tractable by introducing a cutoff, but to check the validity of this artifice, the second-order calculation is carried out entirely in terms of renormalized vertex and propagator functions, in which no cutoff appears.

I. INTRODUCTION

IT has recently been shown[1,2] that the axial-vector current in spinor electrodynamics does not satisfy the usual divergence equation

$$\partial^\mu j_\mu{}^5(x) = 2im_0 j^5(x)\,,$$
$$j_\mu{}^5(x) = \bar\psi(x)\gamma_\mu\gamma_5\psi(x)\,, \quad j^5(x) = \bar\psi(x)\gamma_5\psi(x)\,, \tag{1}$$

expected from naive use of the equations of motion, but rather obeys the equation

$$\partial^\mu j_\mu{}^5(x) = 2im_0 j^5(x) + (\alpha_0/4\pi)F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho}\,, \tag{2}$$

with $F^{\xi\sigma}$ the electromagnetic field strength tensor. Similarly, it was shown that in a simple version of the Gell-Mann–Lévy σ model[3] coupled to the electromagnetic field, the axial-vector current does not satisfy the usual PCAC (partially conserved axial-vector current) equation

$$\partial^\mu j_\mu{}^5(x) = -(\mu_1{}^2/g_0)\pi(x)\,,$$
$$j_\mu{}^5(x) = \bar\psi(x)\tfrac{1}{2}\gamma_\mu\gamma_5\psi(x) + \sigma(x)\partial_\mu\pi(x) - \pi(x)\partial_\mu\sigma(x) + g_0{}^{-1}\partial_\mu\pi(x)\,, \tag{3}$$

but instead obeys the modified PCAC condition

$$\partial^\mu j_\mu{}^5(x) = -(\mu_1{}^2/g_0)\pi(x) + \tfrac{1}{2}(\alpha_0/4\pi)F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho}\,. \tag{4}$$

In both theories, the extra term in Eqs. (2) and (4) arises from the presence of the axial-vector triangle diagram shown in Fig. 1. This diagram has an anomalous property; when it is defined to be gauge-invariant with respect to its vector indices, it does not satisfy the usual axial-vector Ward identity.

An essential conclusion[4] of I was that Eqs. (2) and (4) are exact. In other words, the anomalous term $F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$ does not receive additional contributions from radiative corrections to triangles, such as shown in Fig. 2 (the wavy line denotes either a photon or a meson). This conclusion follows naively from the observation that radiative corrections to the basic triangle involve axial-vector loops (such as the five-vertex loop shown in Fig. 2) which, unlike the lowest-order axial-vector triangle, *do* satisfy the usual axial-vector Ward identity. The purpose of the present paper is to support this naive reasoning with more detailed calculations, and in particular, to show that the fact that radiative corrections to triangles involve the usual renormalizable infinities causes no trouble.

The plan of the paper is as follows. In Sec. II we consider the two models discussed in I—spinor electrodynamics and the σ model—and develop a general argument which shows that Eqs. (2) and (4) are exact. In Sec. III we give an *explicit calculation* of the second-order radiative corrections to the triangle. We find, in agreement with our general arguments, that when all of the second-order radiative corrections are summed, their contributions to the $F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$ term exactly cancel. In Sec. IV we briefly summarize our results and compare them with the conclusions reached recently by Jackiw and Johnson.[2]

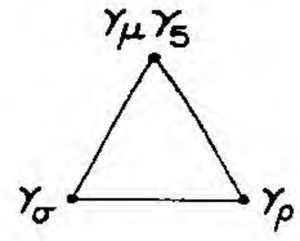


FIG. 1. Axial-vector triangle diagram which leads to the extra term in Eqs. (2) and (4).

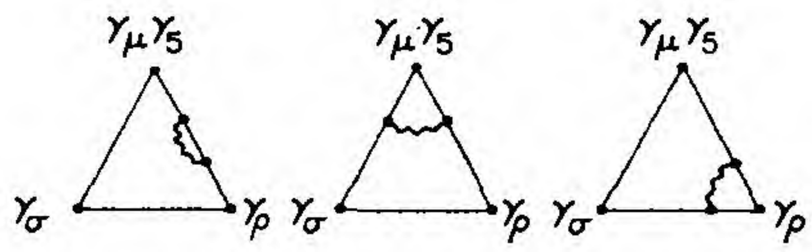


FIG. 2. Typical second-order radiative corrections to the triangle diagram in spinor electrodynamics.

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, U. S. Air Force, under AFOSR Grant No. 68-1365.

[1] S. L. Adler, Phys. Rev. 177, 2426 (1969), hereafter referred to as I. As in I, we use the notation and metric conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill Book Co., New York, 1965), pp. 377–390.

[2] See also J. Schwinger, Phys. Rev. 82, 664 (1951), Sec. V; C. R. Hagen, *ibid.* 177, 2622 (1969); R. Jackiw and K. Johnson, *ibid.* 182, 1459 (1969); R. A. Brandt, *ibid.* 180, 1490 (1969); B. Zumino (unpublished).

[3] M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960). We actually study a truncated version of the σ model proposed by J. S. Bell and R. Jackiw, *ibid.* 60A, 47 (1969).

[4] For example, in order for the low-energy theorem for $\pi^0 \to 2\gamma$ derived in I to be valid, it is essential that there be no strong-interaction corrections to the anomalous term in Eq. (4).

II. GENERAL ARGUMENT

We develop in this section a general argument, valid to any finite order of perturbation theory, which shows that Eqs. (2) and (4) are exact. The basic idea is this: Since Eqs. (2) and (4) involve the *unrenormalized* fields, masses, and coupling constants, these equations are well defined only in a cutoff field theory. Consequently, for both of the field-theoretic models discussed, we construct a cutoff version by introducing photon or meson regulator fields with mass Λ. In both cases, the cutoff prescription allows the usual renormalization program to be carried out, so that the bare masses and couplings and the wave-function renormalizations are specified functions of the renormalized couplings and masses, and of the cutoff Λ. In the cutoff field theories, it is straightforward to prove the validity of Eqs. (2) and (4) for the unrenormalized quantities; this is our principal result. From Eqs. (2) and (4) we obtain exact low-energy theorems for the matrix elements $\langle 2\gamma|2im_0j^5|0\rangle$ and $\langle 2\gamma|(-\mu_1^2/g_0)\pi|0\rangle$; the latter of these yields the $\pi^0\to 2\gamma$ low-energy theorem discussed in I.

Having summarized, in a very condensed way, our method and results, we now turn to the details in the various models.

A. Spinor Electrodynamics

We consider first the case of spinor electrodynamics, described by the Lagrangian density

$$\mathcal{L}(x)=\bar\psi(x)(i\partial\!\!\!/-m_0)\psi(x)-\tfrac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x)-e_0\bar\psi(x)\gamma_\mu\psi(x)A^\mu(x)\,, \quad (5)$$

$$F_{\mu\nu}(x)=\partial_\nu A_\mu(x)-\partial_\mu A_\nu(x)\,,\quad \partial\!\!\!/\equiv\gamma^\mu\partial_\mu\,.$$

We introduce a cutoff by modifying the usual Feynman rules for electrodynamics as follows.

(i) For each internal fermion line with momentum p we include a factor $i(p\!\!\!/-m_0+i\epsilon)^{-1}$ and for each vertex a factor $-ie_0\gamma_\mu$, with m_0 and e_0 the bare mass and charge. For each internal photon line of momentum q, we replace the usual propagator $-ig_{\mu\nu}(q^2+i\epsilon)^{-1}$ by the regulated propagator

$$-ig_{\mu\nu}\left(\frac{1}{q^2+i\epsilon}-\frac{1}{q^2-\Lambda^2+i\epsilon}\right)=\frac{-ig_{\mu\nu}}{q^2+i\epsilon}\,\frac{-\Lambda^2}{q^2-\Lambda^2+i\epsilon}\,. \quad (6)$$

(ii) Let $\Pi_{\mu\nu}^{(2)}(q)$ denote the two-vertex vacuum polarization loop [Fig. 3(a)]

$$\Pi_{\mu\nu}^{(2)}(q)=i\int\frac{d^4k}{(2\pi)^4}\mathrm{Tr}\left[\gamma_\mu\frac{1}{k\!\!\!/-m_0+i\epsilon}\gamma_\nu\frac{1}{k\!\!\!/+q\!\!\!/-m_0+i\epsilon}\right]. \quad (7)$$

Wherever $\Pi_{\mu\nu}^{(2)}(q)$ appears, we use its gauge-invariant, subtracted evaluation[5]

$$\Pi_{\mu\nu}^{(2)}(q)=(q_\mu q_\nu-g_{\mu\nu}q^2)\Pi^{(2)}(q^2)\,,\quad \Pi^{(2)}(0)=0\,. \quad (8)$$

All vacuum polarization loops with four or more vertices [Fig. 3(b)] are calculated by imposing gauge invariance; this suffices to make them finite without need for further subtractions.

(iii) As usual, there is a factor $\int d^4l/(2\pi)^4$ for each internal integration over loop variable l and a factor -1 for each fermion loop.

(iv) We use the standard, iterative renormalization procedure[5] to fix the unrenormalized charge and mass e_0 and m_0 and the fermion wave-function renormalization Z_2 as functions of the renormalized charge and mass e and m and the cutoff Λ. For finite Λ, the renormalization constants e_0, m_0, and Z_2 will all be *finite*. The reason is that regulating the photon propagator (plus gauge invariance for loops) renders finite all vertex and electron self-energy parts and all photon self-energy parts other than $\Pi_{\mu\nu}^{(2)}$. [Examples of such vertex and self-energy parts are given in Fig. 3(c).] The self-energy part $\Pi_{\mu\nu}^{(2)}$ has already been made finite by explicit subtraction.

(v) We include wave-function renormalization factors $Z_2^{1/2}$ and $Z_3^{1/2}$ for each external fermion and photon line. (We recall that $Z_3=e^2/e_0^2$.)

This simple set of rules makes all ordinary electrodynamics matrix elements finite. We may summarize the rules compactly by noting that they are the Feynman rules for the following *regulated* Lagrangian density:

$$\begin{aligned}\mathcal{L}^R(x)&=\mathcal{L}_0^R(x)+\mathcal{L}_I^R(x)\,,\\ \mathcal{L}_0^R(x)&=\bar\psi(x)(i\partial\!\!\!/-m_0)\psi(x)-\tfrac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x)\\ &\quad+\tfrac{1}{4}F_{\mu\nu}{}^R(x)F^{R\mu\nu}(x)-\tfrac{1}{2}\Lambda^2A_\mu{}^R(x)A^{R\mu}(x)\,, \quad (9)\\ \mathcal{L}_I^R(x)&=-e_0\bar\psi(x)\gamma_\mu\psi(x)[A^\mu(x)+A^{R\mu}(x)]\\ &\quad+C^{(2)}[F_{\mu\nu}(x)+F_{\mu\nu}{}^R(x)][F^{\mu\nu}(x)+F^{R\mu\nu}(x)]\,,\end{aligned}$$

where $A_\mu{}^R$ is the field of the regulator vector meson of mass Λ, and $F_{\mu\nu}{}^R(x)=\partial_\nu A_\mu{}^R(x)-\partial_\mu A_\nu{}^R(x)$ is the regulator field strength tensor. The regulator field free

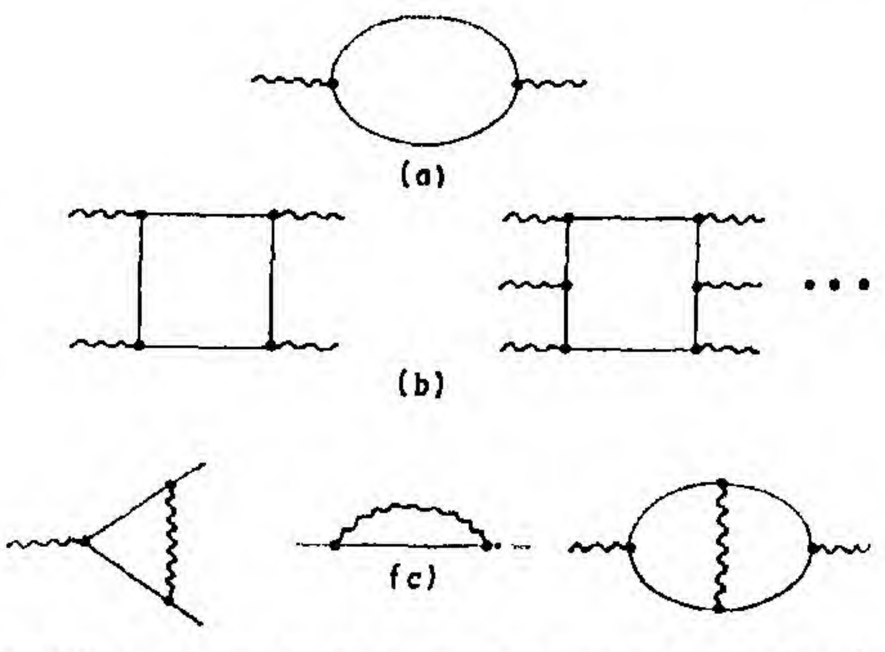


FIG. 3. (a) Two-vertex photon vacuum polarization loop. (b) Larger vacuum polarization loops. (c) Vertex and self-energy parts which are made finite by the photon propagator cutoff and gauge invariance of loops.

[5] Bjorken and Drell (Ref. 1), Chap. 19.

Lagrangian density is included in $\mathcal{L}_0{}^R(x)$ with the opposite sign from normal; hence according to the canonical formalism, the regulator field is quantized with the opposite sign from normal—that is,

$$[A_\mu{}^R(x),\dot{A}_\nu{}^R(y)]|_{x^0=y^0}=ig_{\mu\nu}\delta^3(\mathbf{x}-\mathbf{y}),$$

whereas

$$[A_\mu(x),\dot{A}_\nu(y)]|_{x^0=y^0}=-ig_{\mu\nu}\delta^3(\mathbf{x}-\mathbf{y})$$

—giving the regulator bare propagator the opposite sign from the photon bare propagator. The interaction terms in $\mathcal{L}_I{}^R(x)$ treat the regulator and the photon fields symmetrically. The term proportional to $C^{(2)}$ is a logarithmically infinite counter term which performs the explicit subtraction in the two-vertex vacuum polarization loop $\Pi_{\mu\nu}{}^{(2)}$, so that $\Pi^{(2)}(0)=0$.

Having specified our cutoff procedure, we are now ready to introduce the axial-vector and pseudoscalar currents $j_\mu{}^5(x)$ and $j^5(x)$, and to study their properties. First, we must check whether all matrix elements of these currents are finite when calculated in our cutoff theory. The answer is *yes*, that they are finite, and

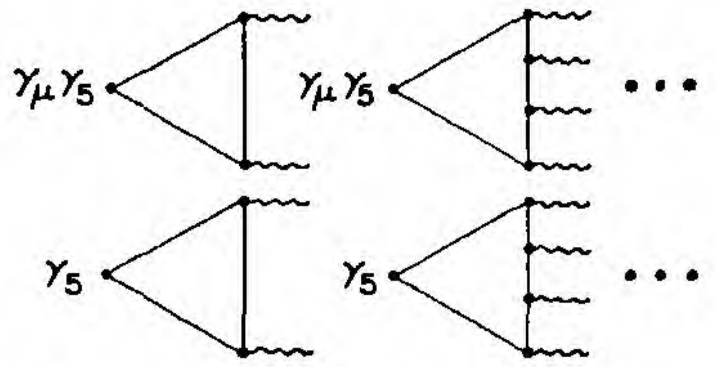


FIG. 4. Basic loops involving one axial-vector or one pseudoscalar vertex.

follows immediately from the fact that all of the basic fermion loops involving one axial-vector or one pseudoscalar vertex (Fig. 4) are made finite by the imposition of gauge invariance on the photon vertices, without the need for explicit subtractions. Thus, we can turn immediately to the problem of showing that Eq. (2) is exactly satisfied in our cutoff theory.

Let us consider an arbitrary Feynman amplitude involving $j_\mu{}^5$, with $2f$ external fermion and b external boson lines (Fig. 5). Proceeding as in I, we divide the diagrams contributing to the Feynman amplitude into two categories, which we call type (a) and type (b). The type-(a) diagrams are those in which the axial-vector vertex $\gamma_\mu\gamma_5$ is attached to one of the f fermion lines running through the diagram; a typical type-(a) contribution is shown in Fig. 6(a). By contrast, the type-(b) diagrams are those in which the axial-vector vertex $\gamma_\mu\gamma_5$ is attached to an internal closed loop; in Fig. 6(b) we show a typical contribution of type (b). In both Figs. 6(a) and 6(b), we have denoted by Q the four-momentum carried by the axial-vector current.

To study the divergence of the axial-vector current, we multiply the matrix element of $j_\mu{}^5$ by iQ^μ. We turn

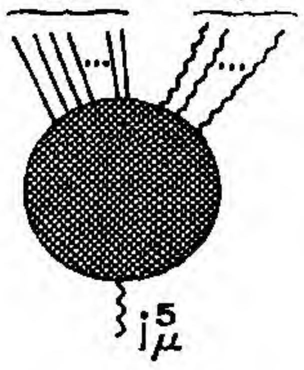


FIG. 5. Arbitrary Feynman amplitude involving $j_\mu{}^5$.

our attention first to the type-(a) contribution pictured in Fig. 6(a), which can be written

$$\sum_{k=1}^{2n-1}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{\not{p}+\not{p}_j-m_0}\right]\gamma^{(k)}\frac{1}{\not{p}+\not{p}_k-m_0}\gamma_\mu\gamma_5\frac{1}{\not{p}'+\not{p}_k-m_0}$$
$$\times\prod_{j=k+1}^{2n-1}\left[\gamma^{(j)}\frac{1}{\not{p}'+\not{p}_j-m_0}\right]\gamma^{(2n)}(\cdots),\quad(10)$$
$$Q=p-p',$$

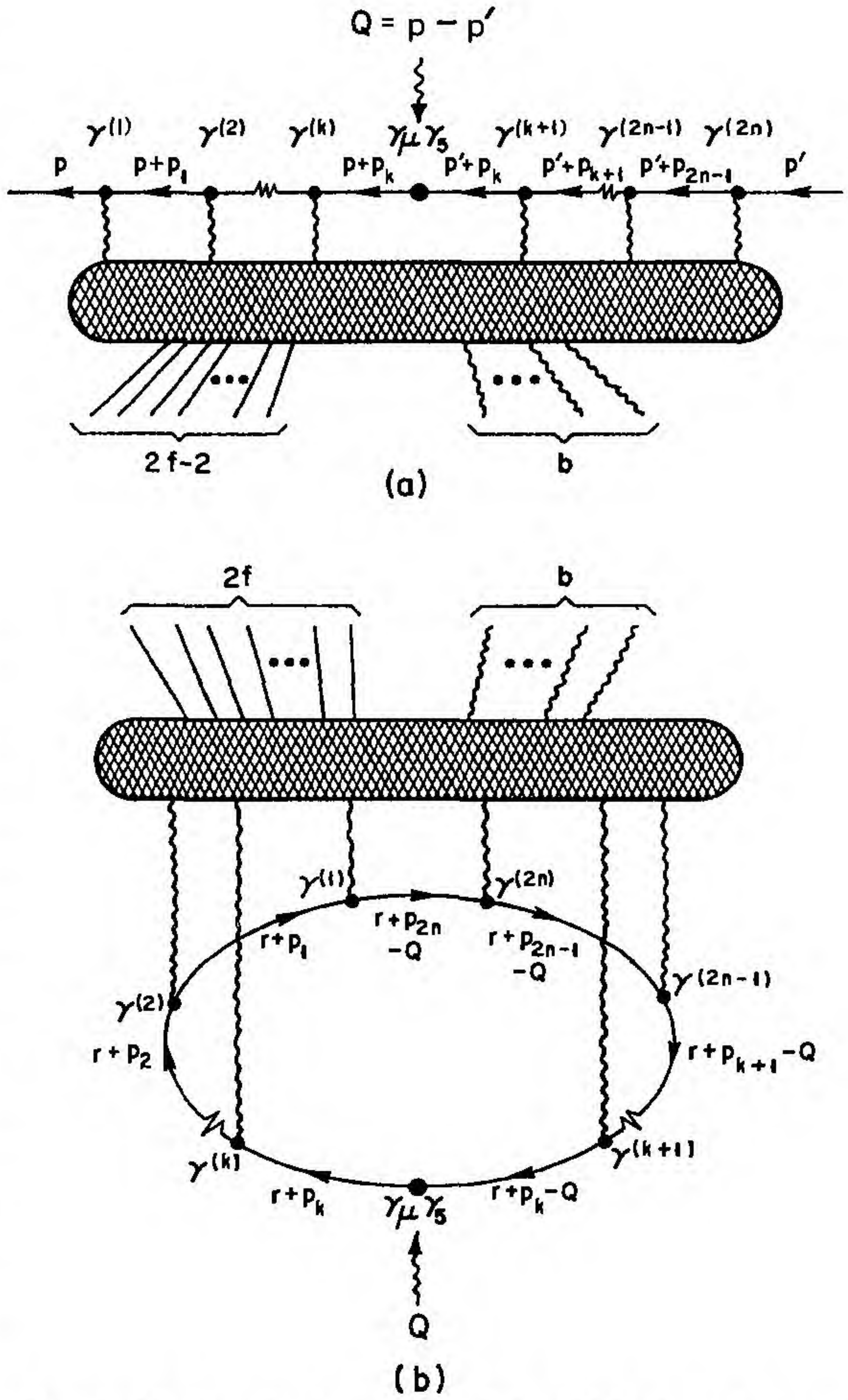


FIG. 6. (a) Contribution to Fig. 5 in which the axial-vector vertex is attached to one of the f fermion lines running through the diagram. (b) Contribution to Fig. 5 in which the axial-vector vertex is attached to an internal closed loop.

where we have focused our attention on the line to which the $\gamma_\mu\gamma_5$ vertex is attached and have denoted the remainder of the diagram by $(\cdots)$. As shown in I, when we multiply the propagator string in Eq. (10) by iQ^μ and do an algebraic rearrangement of terms, we obtain the following identity:

$$iQ^\mu \sum_{k=1}^{2n-1} \prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(k)}\frac{1}{p+p_k-m_0}\gamma_\mu\gamma_5\frac{1}{p'+p_k-m_0}\prod_{j=k+1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}$$
$$=\sum_{k=1}^{2n-1} \prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(k)}\frac{1}{p+p_k-m_0}2im_0\gamma_5\frac{1}{p'+p_k-m_0}\prod_{j=k+1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}$$
$$-i\prod_{j=1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p+p_j-m_0}\right]\gamma^{(2n)}\gamma_5-i\gamma_5\prod_{j=1}^{2n-1}\left[\gamma^{(j)}\frac{1}{p'+p_j-m_0}\right]\gamma^{(2n)}. \quad (11)$$

Since the integrals over the four-momenta of the photon propagators joining the fermion propagator string to the "blob" in Fig. 6(a) (i.e., the integrals over $p_1,\cdots,p_{2n-1}$) are all *convergent* in our regulated field theory, it is safe to do the algebraic manipulations implicit in Eq. (11) *inside* the integrals. The first term on the right-hand side of Eq. (11) gives the type-(a) contribution to the Feynman amplitude for $2im_0j^5$, corresponding to replacing $\gamma_\mu\gamma_5$ by $2im_0\gamma_5$ in Fig. 6(a). The two remaining terms in Eq. (11) are the usual "surface terms" which arise in Ward identities from the equal-time commutator of $j_0{}^5$ with the fields of the external fermions of momenta p and p'.[6] Thus, as far as the type-(a) contributions to the Feynman amplitude are concerned, the divergence of $j_\mu{}^5$ is simply $2im_0j^5$, with no extra terms present.

We turn next to the type-(b) contribution pictured in Fig. 6(b), which we write as

$$L(Q;\gamma_\mu\gamma_5;p_1,\cdots,p_{2n-1})(\cdots),$$

$$L(Q;\Gamma;p_1,\cdots,p_{2n-1})$$
$$=\int d^4r\,\mathrm{Tr}\left\{\sum_{k=1}^{2n}\prod_{j=1}^{k-1}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\right.$$
$$\times\gamma^{(k)}\frac{1}{r+p_k-m_0}\Gamma\frac{1}{r+p_k-Q-m_0}$$
$$\left.\times\prod_{j=k+1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j-Q-m_0}\right]\right\}, \quad (12)$$

where we have focused our attention on the closed loop and have again denoted the remainder of the diagram by $(\cdots)$. As was shown in I, some straightforward algebra implies

$$L(Q;iQ^\mu\gamma_\mu\gamma_5;p_1,\cdots,p_{2n-1})$$
$$=L(Q;2im_0\gamma_5;p_1,\cdots,p_{2n-1})$$
$$+i\int d^4r\,\mathrm{Tr}\left\{\gamma_5\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j-m_0}\right]\right.$$
$$\left.-\gamma_5\prod_{j=1}^{2n}\left[\gamma^{(j)}\frac{1}{r+p_j-Q-m_0}\right]\right\}. \quad (13)$$

For loops with $n\geq 2$ (i.e., with four or more vector vertices) the residual integral in Eq. (13) is sufficiently convergent for us to be able to make the change of variable $r\to r+Q$ in the second term, causing the two terms in the curly brackets to cancel. Thus, the loops with $n\geq 2$ satisfy the usual Ward identity

$$L(Q;iQ^\mu\gamma_\mu\gamma_5;p_1,\cdots,p_{2n-1})$$
$$=L(Q;2im_0\gamma_5;p_1,\cdots,p_{2n-1}). \quad (14)$$

Again, since the integrals over $p_1,\cdots,p_{2n-1}$ are all *convergent* in the regulated field theory, the manipulations leading to Eq. (14) can all be performed *inside* these integrals. This means that the type-(b) pieces containing loops with $n\geq 2$ all agree with the usual divergence equation $\partial^\mu j_\mu{}^5(x)=2im_0j^5(x)$.

Finally, we must consider the case $n=1$, that is, the axial-vector triangle diagram illustrated in Fig. 7. As was shown in I, when the triangle is defined to be gauge-invariant with respect to the vector indices, it does *not* satisfy Eq. (14) for the axial-vector index divergence. Instead, there is a well-defined extra term left over which comes from the failure of the two terms in the curly brackets in Eq. (13) to cancel. The analysis of I shows that the effect of the extra term is to add to the normal axial-vector divergence equation the term

$$(\alpha_0/4\pi)[F^{\xi\sigma}(x)+F^{R\xi\sigma}(x)]$$
$$\times[F^{\tau\rho}(x)+F^{R\tau\rho}(x)]\epsilon_{\xi\sigma\tau\rho}. \quad (15)$$

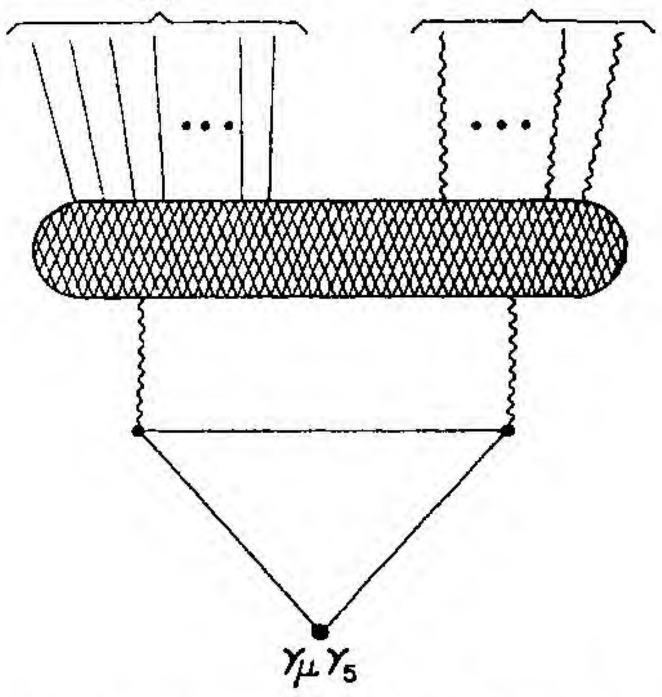


FIG. 7. Contribution of the axial-vector triangle diagram to the Feynman amplitude of Fig. 5.

[6] Y. Takahashi, Nuovo Cimento 6, 370 (1957).

To summarize, our diagrammatic analysis shows that the axial-vector divergence equation in the regulated field theory is

$$\begin{aligned}\partial^\mu j_\mu{}^5(x) = 2im_0 j^5(x) &+ (\alpha_0/4\pi)F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho} \\ &+ (\alpha_0/4\pi)[F^{\xi\sigma}(x)F^{R\tau\rho}(x) + F^{R\xi\sigma}(x)F^{\tau\rho}(x) \\ &\qquad + F^{R\xi\sigma}(x)F^{R\tau\rho}(x)]\epsilon_{\xi\sigma\tau\rho}. \qquad (16)\end{aligned}$$

Equation (16) is identical with Eq. (2), apart from the terms involving F^R which arise from our explicit inclusion of a regulator field. The crucial point is that the coefficient of the anomalous term is exactly $\alpha_0/4\pi$ and *does not involve an unknown power series in the coupling constant coming from higher orders in perturbation theory.*

The diagrammatic analysis which we have just given may be rephrased succinctly as follows: If we use the regulated Lagrangian density in Eq. (9) to calculate equations of motion, and then use the equations of motion to naively calculate the axial divergence, we find

$$\partial^\mu j_\mu{}^5(x) = 2im_0 j^5(x). \qquad (17)$$

Extra terms on the right-hand side of Eq. (17) can only come from singular diagrams where the naive derivation breaks down. In the regulated field theory, all virtual photon integrations *converge* and therefore, cannot lead to singularities giving additional terms in Eq. (17). Hence breakdown in Eq. (17) (if it occurs at all) must be associated with the basic axial-vector loops shown in Fig. 4. But, as we have seen, the axial-vector loops with four or more photons satisfy Eq. (17), so the basic triangle diagram is the *only* possible source of an anomaly.

Having derived our basic result, we turn next to the low-energy theorem for $2im_0 j^5(x)$ which is implied by Eq. (16). Taking the matrix element of Eq. (16) between a state with two photons and the vacuum gives

$$F(k_1\cdot k_2) = G(k_1\cdot k_2) + H(k_1\cdot k_2), \qquad (18)$$

where

$$\begin{aligned}&\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|\partial^\mu j_\mu{}^5|0\rangle \\ &\quad = (4k_{10}k_{20})^{-1/2}k_1{}^\xi k_2{}^\tau \epsilon_1{}^{*\sigma}\epsilon_2{}^{*\rho}\epsilon_{\xi\tau\sigma\rho}F(k_1\cdot k_2), \\ &\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|2im_0 j^5|0\rangle \\ &\quad = (4k_{10}k_{20})^{-1/2}k_1{}^\xi k_2{}^\tau \epsilon_1{}^{*\sigma}\epsilon_2{}^{*\rho}\epsilon_{\xi\tau\sigma\rho}G(k_1\cdot k_2), \\ &(\alpha_0/4\pi)\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|(F^{\xi\sigma}+F^{R\xi\sigma})(F^{\tau\rho}+F^{R\tau\rho})\epsilon_{\xi\sigma\tau\rho}|0\rangle \\ &\quad = (4k_{10}k_{20})^{-1/2}k_1{}^\xi k_2{}^\tau \epsilon_1{}^{*\sigma}\epsilon_2{}^{*\rho}\epsilon_{\xi\tau\sigma\rho}H(k_1\cdot k_2). \qquad (19)\end{aligned}$$

We wish, in particular, to study Eq. (18) at the point $k_1\cdot k_2 = 0$. As has been shown by Sutherland and Veltman,[7] $F(k_1\cdot k_2) \propto k_1\cdot k_2$, so the left-hand side of Eq. (18) vanishes at $k_1\cdot k_2 = 0$, giving

$$G(0) = -H(0). \qquad (20)$$

There are two types of diagrams which contribute to $H(k_1\cdot k_2)$, as illustrated in Fig. 8, where we have used the symbol $\otimes$ to denote action of the operator $(\alpha_0/4\pi)(F^{\xi\sigma}+F^{R\xi\sigma})(F^{\tau\rho}+F^{R\tau\rho})\epsilon_{\xi\sigma\tau\rho}$. In the diagrams in Fig. 8(a), the field strength operators attach directly onto the external photon lines, without photon-photon scattering. The effect of the vacuum polarization parts and the external-line wave-function renormalizations is to change α_0 to α, giving

$$H(0)^{(a)} = 2\alpha/\pi. \qquad (21)$$

In the diagrams in Fig. 8(b), there is a photon-photon scattering between $\otimes$ and the free photons. As a result of the antisymmetric tensor structure of the anomalous divergence term, the vertex $\otimes$ is proportional to k_1+k_2. Also, the diagram for the scattering of light by light is itself proportional to k_1k_2, since photon gauge invariance implies that the external photons couple through their field strength tensors.[8] Thus, the diagrams in Fig. 8(b) are proportional to $k_1k_2(k_1+k_2)$ and are of higher order than the terms which contribute to the low-energy theorem, giving us

$$H(0)^{(b)} = 0. \qquad (22)$$

Combining Eqs. (20)–(22), we get an *exact low-energy theorem* for the operator $2im_0 j^5$,

$$G(0) = -2\alpha/\pi. \qquad (23)$$

So far in our discussion we have kept the cutoff Λ finite, so that $G(0)$ is a matrix element calculated with our modified Feynman rules. Let us now indicate briefly the form which Eq. (23) takes when the cutoff Λ becomes infinite. A straightforward analysis of matrix elements of the operator j^5 shows that divergences as $\Lambda \to \infty$ are associated *only* with the electron propagator

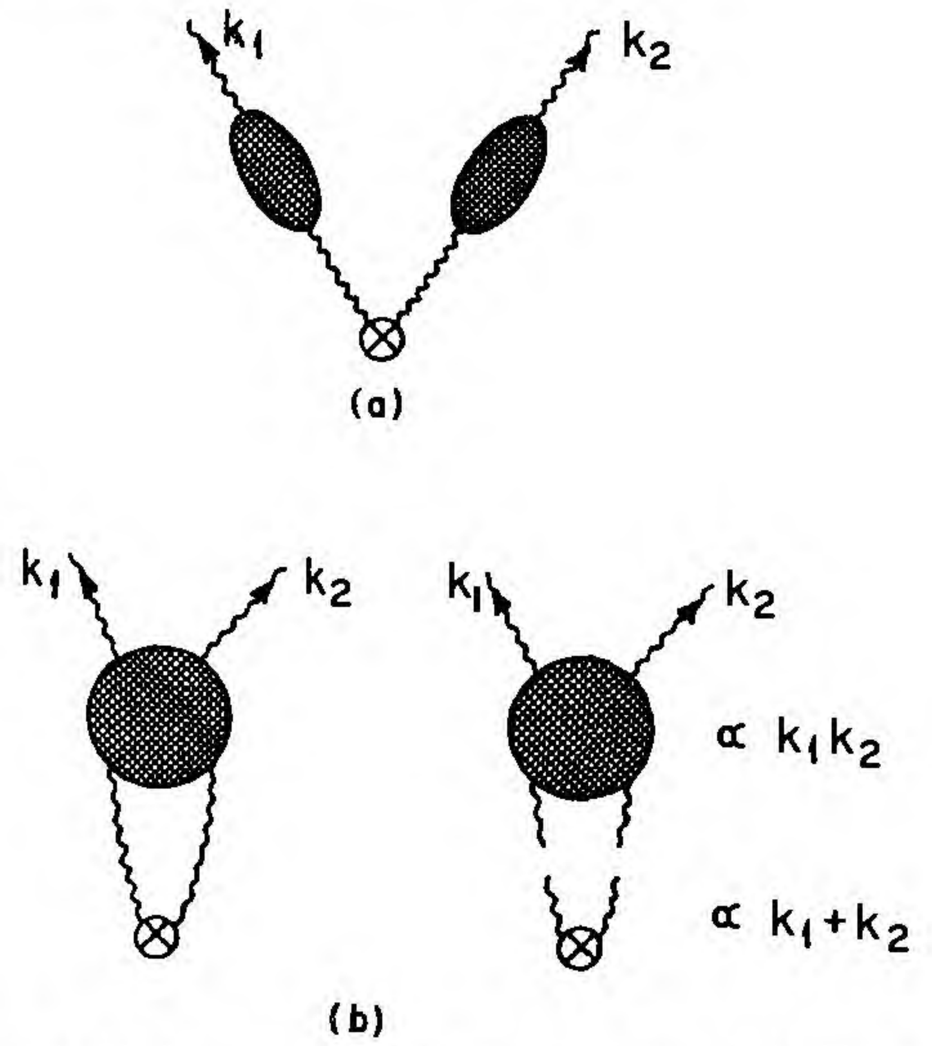


FIG. 8. (a) Diagram in which the operator $(\alpha_0/4\pi)(F^{\xi\sigma}+F^{R\xi\sigma})\times(F^{\tau\rho}+F^{R\tau\rho})\epsilon_{\xi\sigma\tau\rho}$, denoted by $\otimes$, attaches directly to the external photon lines. (b) Diagram in which there is a photon-photon scattering between $\otimes$ and the two external photons.

[7] D. G. Sutherland, Nucl. Phys. B2, 433 (1967).

[8] R. Karplus and M. Neumann, Phys. Rev. 80, 380; 83, 776 (1950).

$S_F'(p)$, the photon propagator $D_F'(q)_{\mu\nu}$, the photon vertex part $\Gamma_\mu(p,p')$, and, in addition, the pseudoscalar vertex part $\Gamma^5(p,p')$, defined by

$$S_F'(p)\Gamma^5(p,p')S_F'(p') = -\int d^4x d^4y\, e^{ip\cdot x}e^{-ip'\cdot y} \times\langle T(\psi(x)j^5(0)\bar{\psi}(y))\rangle_0. \quad (24)$$

In matrix elements of $m_0 j^5$, the vertex part $\Gamma^5(p,p')$ will clearly always occur in the combination $m_0\Gamma^5(p,p')$. Let us now introduce the usual electrodynamic renormalizations

$$\begin{aligned} Z_2^{-1}S_F'(p) &= \tilde{S}_F'(p)\,, \\ Z_3^{-1}D_F'(q)_{\mu\nu} &= \tilde{D}_F'(q)_{\mu\nu}\,, \\ Z_1\Gamma_\mu(p,p') &= \tilde{\Gamma}_\mu(p,p')\,, \\ Z_1 = Z_2\,, \quad e_0 &= Z_3^{-1/2}e\,, \end{aligned} \quad (25)$$

plus the usual wave-function renormalizations on external lines. The effect of these rescalings is to replace $m_0\Gamma^5(p,p')$, wherever it occurs, by $m_0Z_2\Gamma^5(p,p')$. But, as we will now show, this latter quantity is *finite* (i.e., cutoff-independent as $\Lambda\to\infty$). To see this, we note first that $\Gamma^5(p,p')$ is multiplicatively renormalizable[9]; therefore, if $Z_5\Gamma^5(p,p)$ is finite, then so is $Z_5\Gamma^5(p,p')$. Next, let us write the Ward identity for the axial-vector vertex,

$$(p-p')^\mu\Gamma_\mu{}^5(p,p') = 2m_0\Gamma^5(p,p') - i(\alpha_0/4\pi)\tilde{F}(p,p') + S_F'(p)^{-1}\gamma_5 + \gamma_5 S_F'(p')^{-1}\,, \quad (26)$$

where $\Gamma_\mu{}^5$ and $\tilde{F}$ are defined by replacing $j^5(0)$ in Eq. (24) by $j_\mu{}^5(0)$ and $F^{\xi\sigma}(0)F^{\tau\rho}(0)\epsilon_{\xi\sigma\tau\rho}$, respectively. When $p=p'$, the left-hand side of Eq. (26) obviously vanishes. It is also easy to see that $\tilde{F}(p,p)=0$ as a result of the antisymmetric tensor structure of this term. So the axial-vector Ward identity at $p=p'$ becomes the simple equation

$$0 = 2m_0\Gamma^5(p,p) + S_F'(p)^{-1}\gamma_5 + \gamma_5 S_F'(p)^{-1}\,, \quad (27)$$

which immediately implies that $m_0Z_2\Gamma^5(p,p)$ is finite. If we introduce a renormalized pseudoscalar vertex part by writing

$$m_0Z_2\Gamma^5(p,p') = m\tilde{\Gamma}^5(p,p')\,, \quad (28)$$

then Eq. (28) may be rewritten as the equal-momentum boundary condition

$$0 = 2m\tilde{\Gamma}^5(p,p) + \tilde{S}_F'(p)^{-1}\gamma_5 + \gamma_5\tilde{S}_F'(p)^{-1}. \quad (29)$$

The results of this analysis may be summarized as follows: If we calculate an arbitrary matrix element of $m_0 j^5$ in our cutoff theory, and then let $\Lambda\to\infty$, we get the same answer as if we calculated all the skeleton diagrams for the matrix element and replaced the electron and photon lines and the vertex parts appearing in the skeleton by the renormalized quantities $\tilde{S}_F'(p)$, $\tilde{D}_F'(q)_{\mu\nu}$, $\tilde{\Gamma}_\mu(p,p')$, and $m\tilde{\Gamma}^5(p,p')$. These quantities can all be calculated without recourse to cutoffs by using dispersive methods; in the case of the pseudoscalar vertex the subtraction constant in the dispersion relation can be fixed by using Eq. (29) as a boundary condition.

Returning to our low-energy theorem, we see that in the limit $\Lambda\to\infty$ the pseudoscalar-photon-photon matrix element $G(k_1\cdot k_2)$ becomes the renormalized matrix element $\tilde{G}(k_1\cdot k_2)$ calculated by the recipe we have just outlined, and the low-energy theorem tells us that

$$\tilde{G}(0) = -2\alpha/\pi\,. \quad (30)$$

In other words, all order α^2, α^3, $\cdots$, contributions to $\tilde{G}(k_1\cdot k_2)$ vanish at $k_1\cdot k_2=0$. In the next section, we will verify by explicit calculation that the order α^2 terms do cancel.

In conclusion, we remark that the arguments which we have given in this section for spinor electrodynamics apply, with only trivial modification, to the neutral-vector-meson model of strong interactions. In particular the low-energy theorem analogous to Eq. (23) will hold to all orders in both α and the neutral-vector-meson strong coupling.

B. σ Model

We turn next to the case of Gell-Mann and Lévy's σ model.[3] As in I, we consider a truncated version of the σ model which contains only a proton (ψ), a neutral pion (π), and a scalar meson (σ), but omits the charged pions and the neutron. Our exposition will differ somewhat from the previous case of spinor electrodynamics, where we *first* introduced a cutoff procedure to remove infinities from the theory, and then afterwards proceeded to discuss the properties of the axial-vector current. In the case of the σ model, we will have to consider the axial-vector current *simultaneously* with our introduction of the cutoff, in order to ensure that the cutoff preserves the usual PCAC equation [Eq. (3)] when electromagnetic interactions are neglected. Once we are sure that the axial-vector divergence equation in the absence of electromagnetism has no abnormalities, we can then determine how Eq. (3) is modified when electromagnetic effects are taken into account.

We begin by writing the Lagrangian for the σ model and discussing some of the formal properties of this theory. We have

$$\begin{aligned} \mathcal{L} = {} & \bar{\psi}[i\partial - G_0(g_0^{-1}+\sigma+i\pi\gamma_5)]\psi \\ & + \lambda_0[4\sigma^2+4g_0\sigma(\sigma^2+\pi^2)+g_0^2(\sigma^2+\pi^2)^2] \\ & + \tfrac{1}{2}\mu_0^2[2g_0^{-1}\sigma+\sigma^2+\pi^2] \\ & + \tfrac{1}{2}[(\partial\pi)^2+(\partial\sigma)^2] - \tfrac{1}{2}\mu_1^2(\pi^2+\sigma^2)\,, \end{aligned} \quad (31)$$

[9] We recall that multiplicative renormalizability of the usual vector vertex Γ_ρ follows from the fact that Γ_ρ satisfies the integral equation $\Gamma_\rho = \gamma_\rho - \int\Gamma_\rho S_F' S_F' K$ [see Bjorken and Drell (Ref. 5)]. More generally, G. Preparata and W. I. Weisberger [Phys. Rev. 175, 1965 (1968)] have observed that in spinor electrodynamics the vertex $\Gamma_O(p,p')$ of $\bar{\psi}O\psi$, with O a product of γ-matrices, satisfies the integral equation $\Gamma_O = O - \int\Gamma_O S_F' S_F' K$, and therefore is multiplicatively renormalizable.

where we have chosen the fully translated form of the σ model with

$$\langle\sigma\rangle_0=0 \tag{32}$$

to all orders of perturbation theory. The axial-vector current is generated by the chiral gauge transformation

$$\begin{aligned} \psi &\to (1+\tfrac{1}{2}i\gamma_5 v)\psi\,, \\ \pi &\to \pi - v(g_0^{-1}+\sigma)\,, \\ g_0^{-1}+\sigma &\to g_0^{-1}+\sigma+v\pi\,, \end{aligned} \tag{33}$$

giving

$$\begin{aligned} j_\mu{}^5 &= -\delta\mathfrak{L}/\delta(\partial^\mu v) = \bar{\psi}\tfrac{1}{2}\gamma_\mu\gamma_5\psi+\sigma\partial_\mu\pi-\pi\partial_\mu\sigma+g_0^{-1}\partial_\mu\pi\,, \\ \partial^\mu j_\mu{}^5 &= -\delta\mathfrak{L}/\delta v = -(\mu_1{}^2/g_0)\pi\,. \end{aligned} \tag{34}$$

The terms in Eq. (31) have the following significance: (i) G_0 is the unrenormalized meson-nucleon coupling constant; (ii) the quantity g_0 is related to the bare nucleon mass m_0 by

$$G_0/g_0=m_0\,, \tag{35}$$

and may be expressed directly as a vacuum expectation value,

$$1/g_0=i\left\langle\left[\int d^3x\, j_0{}^5(x),\pi(0)\right]\Big|_{x_0=0}\right\rangle_0; \tag{36}$$

(iii) $\mu_1{}^2$ is the bare meson mass which appears in the bare propagators $\Delta_F{}^\sigma(q)$ and $\Delta_F{}^\pi(q)$,

$$\Delta_F{}^{\sigma,\pi}(q)=1/(q^2-\mu_1{}^2+i\epsilon); \tag{37}$$

(iv) the term $\lambda_0[4\sigma^2+4g_0\sigma(\sigma^2+\pi^2)+g_0{}^2(\sigma^2+\pi^2)^2]$ is a chiral-invariant meson-meson scattering interaction; and (v) the term $\frac{1}{2}\mu_0{}^2[2g_0{}^{-1}\sigma+\sigma^2+\pi^2]$ is a chiral-invariant *counter term* which is necessary to guarantee that

$$\langle\delta\mathfrak{L}/\delta\sigma\rangle_0=\partial_\lambda\langle\delta\mathfrak{L}/\delta(\partial_\lambda\sigma)\rangle_0=0\,, \tag{38}$$

as is required by the Euler-Lagrange equations of motion and translation invariance. Equations (32) and (38) fix $\mu_0{}^2$ to have the value

$$\mu_0{}^2=\langle G_0g_0\bar{\psi}\psi-\lambda_0[4g_0{}^2(3\sigma^2+\pi^2)+4g_0{}^3\sigma(\sigma^2+\pi^2)]\rangle_0\,. \tag{39}$$

The effect of $\mu_0{}^2$, which is formally quadratically divergent, is to remove the "tadpole" diagrams of the type shown in Fig. 9, so that the condition $\langle\sigma\rangle_0=0$ is maintained in each order of perturbation theory. It is easily seen that the $\mu_0{}^2$ counter term simultaneously removes the quadratically divergent parts of the π- and σ-meson self-energies, so that the remaining bare quantities appearing in the Lagrangian (G_0,g_0,μ_1), as well as the wave-function renormalizations, are at most *logarithmically* divergent.

FIG. 9. Tadpole diagram.

An important feature of the Lagrangian density in Eq. (31) is that it is *not* normal-ordered; the omission of normal ordering is essential in order for the axial-vector current to satisfy the PCAC equation (34). To see this, let us consider the effect of normal ordering on the chiral-invariant meson-meson scattering term,

$$\begin{aligned} &\mathfrak{L}_{MM}=\mathfrak{L}_{MM}{}^{(2)}+\mathfrak{L}_{MM}{}^{(3)}+\mathfrak{L}_{MM}{}^{(4)}\,, \\ &\mathfrak{L}_{MM}{}^{(2)}=4\sigma^2\,,\quad \mathfrak{L}_{MM}{}^{(3)}=4g_0\sigma(\sigma^2+\pi^2)\,, \\ &\mathfrak{L}_{MM}{}^{(4)}=g_0{}^2(\sigma^2+\pi^2)^2\,. \end{aligned} \tag{40}$$

The normal-ordered forms of the two-, three-, and four-meson scattering terms are defined by

$$\begin{aligned} \langle:\mathfrak{L}_{MM}{}^{(2)}:\rangle_0 &= \langle(\partial/\partial\sigma):\mathfrak{L}_{MM}{}^{(2)}:\rangle=0\,, \\ \langle:\mathfrak{L}_{MM}{}^{(3)}:\rangle_0 &= \langle(\partial/\partial\sigma):\mathfrak{L}_{MM}{}^{(3)}:\rangle_0=\langle(\partial/\partial\pi):\mathfrak{L}_{MM}{}^{(3)}:\rangle_0 \\ &= \langle(\partial^2/\partial\sigma^2):\mathfrak{L}_{MM}{}^{(3)}:\rangle_0 \\ &= \langle(\partial/\partial\sigma)(\partial/\partial\pi):\mathfrak{L}_{MM}{}^{(3)}:\rangle_0 \\ &= \langle(\partial^2/\partial\pi^2):\mathfrak{L}_{MM}{}^{(3)}:\rangle_0=0\,, \\ \langle:\mathfrak{L}_{MM}{}^{(4)}:\rangle_0 &= \langle(\partial/\partial\sigma):\mathfrak{L}_{MM}{}^{(4)}:\rangle_0=\langle(\partial/\partial\pi):\mathfrak{L}_{MM}{}^{(4)}:\rangle_0 \\ &= \langle(\partial^2/\partial\sigma^2):\mathfrak{L}_{MM}{}^{(4)}:\rangle_0=\cdots \\ &= \langle(\partial^3/\partial\pi^3):\mathfrak{L}_{MM}{}^{(4)}:\rangle_0=0\,. \end{aligned} \tag{41}$$

These conditions may easily be satisfied by introducing counter terms to remove the vacuum expectation values of the various derivatives,

$$\begin{aligned} :\mathfrak{L}_{MM}{}^{(2)}: &= 4\sigma^2-\langle4\sigma^2\rangle_0\,, \\ :\mathfrak{L}_{MM}{}^{(3)}: &= 4g_0\sigma(\sigma^2+\pi^2)-4g_0\sigma\langle3\sigma^2+\pi^2\rangle_0 \\ &\quad -4g_0\langle\sigma(\sigma^2+\pi^2)\rangle_0\,, \\ :\mathfrak{L}_{MM}{}^{(4)}: &= g_0{}^2(\sigma^2+\pi^2)^2-4g_0{}^2\sigma\langle\sigma(\sigma^2+\pi^2)\rangle_0 \\ &\quad -2g_0{}^2\sigma^2\langle3\sigma^2+\pi^2\rangle_0-2g_0{}^2\pi^2\langle\sigma^2+3\pi^2\rangle_0 \\ &\quad -g_0{}^2\langle(\sigma^2+\pi^2)^2\rangle_0+2g_0{}^2\langle\sigma^2\rangle_0\langle3\sigma^2+\pi^2\rangle_0 \\ &\quad +2g_0{}^2\langle\pi^2\rangle_0\langle\sigma^2+3\pi^2\rangle_0\,, \end{aligned} \tag{42}$$

giving

$$\begin{aligned} :\mathfrak{L}_{MM}: &= \mathfrak{L}_{MM}-4g_0\sigma\langle3\sigma^2+\pi^2\rangle_0-4g_0{}^2\sigma\langle\sigma(\sigma^2+\pi^2)\rangle_0 \\ &\quad -2g_0{}^2\sigma^2\langle3\sigma^2+\pi^2\rangle_0-2g_0{}^2\pi^2\langle\sigma^2+3\pi^2\rangle_0 \\ &\quad +\text{const}\,. \end{aligned} \tag{43}$$

Clearly, the normal-ordered interaction $:\mathfrak{L}_{MM}:$ will be chiral-invariant only if the counter terms combine to be proportional to $\sigma^2+\pi^2+2\sigma/g_0$, that is, only if

$$\langle3\sigma^2+\pi^2\rangle_0=\langle\sigma^2+3\pi^2\rangle_0=\langle3\sigma^2+\pi^2\rangle_0+g_0\langle\sigma(\sigma^2+\pi^2)\rangle_0\,, \tag{44}$$

which requires

$$\begin{aligned} \langle\sigma^2\rangle_0&=\langle\pi^2\rangle_0\,, \\ \langle\sigma(\sigma^2+\pi^2)\rangle_0&=0\,. \end{aligned} \tag{45}$$

These conditions would be satisfied if π and σ were free fields, but they are *not* true in the presence of the interaction terms of Eq. (31). Thus, the normal-ordered form $:\mathfrak{L}_{MM}:$ is not invariant under the chiral gauge transformation of Eq. (33) and, if used in the Lagrangian instead of $\mathfrak{L}_{MM}$, spoils the PCAC equation. The way out of this difficulty consists in noting that the

normal-ordering conditions of Eq. (4) are not necessary for the consistency of the Lagrangian field theory of Eq. (31); all that is necessary is the single condition $\langle\delta\mathcal{L}/\delta\sigma\rangle_0=0$. As we have seen, this condition can be satisfied by including the chiral-invariant counter term proportional to μ_0^2, without any use of normal ordering in the Lagrangian.

The fact that μ_0^2 removes the quadratic divergence from the π and σ self-energies can be expressed in a simple equation, which will be very useful in what follows. Let $\Delta_F^{\pi\prime}(q)$ denote the full pion propagator, given by

$$\Delta_F^{\pi\prime}(q)=-i\int d^4x\, e^{iq\cdot x}\langle T(\pi(x)\pi(0))\rangle_0$$
$$=1/[q^2-\mu_1^2-\Sigma^\pi(q^2)], \tag{46}$$

where $\Sigma^\pi(q^2)$ is the pion proper self-energy. According to Eq. (46),

$$\Delta_F^{\pi\prime}(0)=-1/[\mu_1^2+\Sigma^\pi(0)]. \tag{47}$$

An alternative expression for $\Delta_F^{\pi\prime}(q)$ may be obtained by substituting the PCAC equation (34) into Eq. (46),

$$\Delta_F^{\pi\prime}(q)=i\frac{g_0}{\mu_1^2}\int d^4x\, e^{iq\cdot x}\langle T((\partial/\partial x_\mu)j_\mu^5(x)\pi(0))\rangle_0$$
$$=\frac{g_0}{\mu_1^2}q^\mu\int d^4x\, e^{iq\cdot x}\langle T(j_\mu^5(x)\pi(0))\rangle_0$$
$$-\frac{ig_0}{\mu_1^2}\int d^3x\, e^{-i\mathbf{q}\cdot\mathbf{x}}\langle[j_0^5(x),\pi(0)]|_{x_0=0}\rangle_0, \tag{48}$$

which at $q=0$ becomes

$$\Delta_F^{\pi\prime}(0)=-\frac{ig_0}{\mu_1^2}\int d^3x\,\langle[j_0^5(x),\pi(0)]|_{x_0=0}\rangle_0=\frac{-1}{\mu_1^2}. \tag{49}$$

Comparing Eqs. (47) and (49), we obtain the desired result

$$\Sigma^\pi(0)=0. \tag{50}$$

Since the differences $\Sigma^\pi(q^2)-\Sigma^\pi(0)$ and $\Sigma^\sigma(q^2)-\Sigma^\pi(0)$ are only logarithmically divergent, Eq. (50) tells us that the π and σ self-energies $\Sigma^\pi(q^2)$ and $\Sigma^\sigma(q^2)$ are themselves only logarithmically divergent.

So far, we have discussed the σ model in the absence of electromagnetism. To include electromagnetism, we add to the Lagrangian density of Eq. (31) the terms

$$-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}-e_0\bar{\psi}\gamma_\mu\psi A^\mu. \tag{51}$$

We expect, because of the presence of triangle diagrams, that electromagnetism will modify the PCAC equation by the addition of a term proportional to $F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$. However, it is easy to see that all of the other formal properties of the σ model which we have derived above are unchanged. In particular, *Eq. (50) is still valid in the presence of electromagnetism*, since the antisymmetric tensor structure of the extra term in the PCAC equation causes the contribution of this term to Eq. (48) to vanish at $q=0$.

This completes our survey of the formal properties of the σ model. We proceed to introduce a cutoff (with electromagnetism included) by modifying the usual Feynman rules as follows.

(i) For each internal fermion line with momentum p we include a factor $i(\not{p}-m_0+i\epsilon)^{-1}$, with $m_0=G_0/g_0$. For each internal photon line of momentum q, we replace the usual propagator $-ig_{\mu\nu}(q^2+i\epsilon)^{-1}$ by the regulated propagator

$$-ig_{\mu\nu}\left(\frac{1}{q^2+i\epsilon}-\frac{1}{q^2-\Lambda^2+i\epsilon}\right)=\frac{-ig_{\mu\nu}}{q^2+i\epsilon}\left(\frac{-\Lambda^2}{q^2-\Lambda^2+i\epsilon}\right). \tag{52}$$

For internal π or σ lines, *which are not attached at either end to the axial-vector current*, we replace the usual propagator $i(q^2-\mu_1^2+i\epsilon)^{-1}$ by the regulated propagator

$$i\left(\frac{1}{q^2-\mu_1^2+i\epsilon}-\frac{1}{q^2-\Lambda^2+i\epsilon}\right)$$
$$=\frac{i}{q^2-\mu_1^2+i\epsilon}\left(\frac{-\Lambda^2+\mu_1^2}{q^2-\Lambda^2+i\epsilon}\right). \tag{53}$$

For the photon-nucleon, meson-nucleon, and meson-meson vertices, we include the factors shown in Fig. 10, with e_0, g_0, G_0, and λ_0 the appropriate bare couplings.

(ii) For the axial-vector-current–nucleon and axial-vector-current–meson vertices, we include the factors shown in Fig. 11. For the pion propagator immediately following the axial-vector-current–pion vertex, we use the *unregulated* propagator $i(q^2-\mu_1^2+i\epsilon)^{-1}$, while we replace the *product* of pion and σ propagators immediately following the axial-vector-current–pion–σ

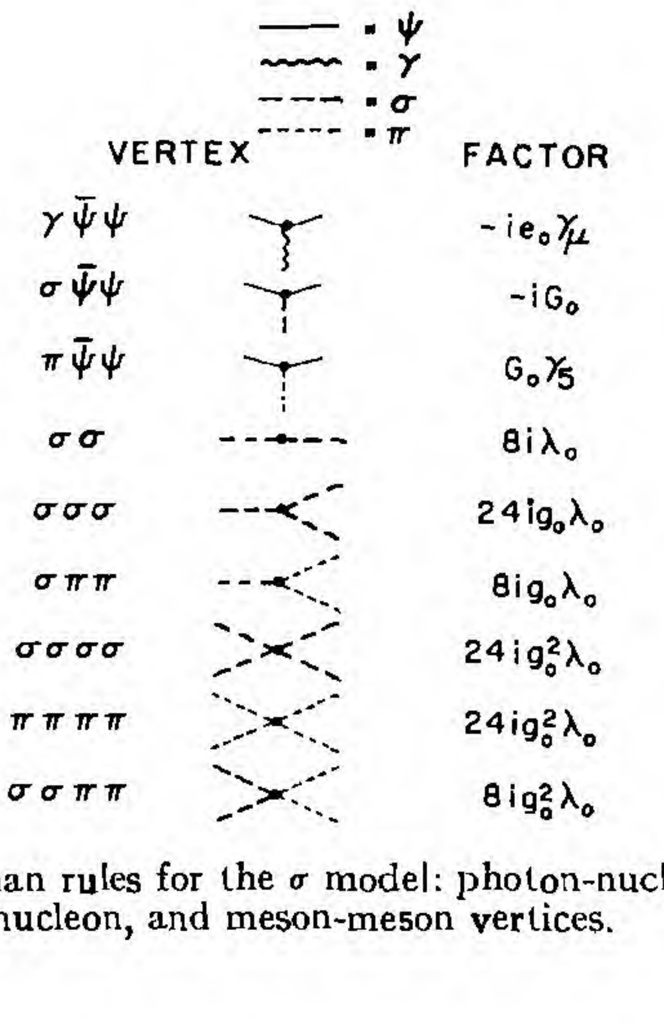


FIG. 10. Feynman rules for the σ model: photon-nucleon, meson-nucleon, and meson-meson vertices.

vertex by

$$\frac{i}{q^2-\mu_1^2+i\epsilon}\,\frac{i}{(q+Q)^2-\mu_1^2+i\epsilon}\,\frac{i}{q^2-\Lambda^2+i\epsilon}\,\frac{i}{(q+Q)^2-\Lambda^2+i\epsilon}. \tag{54}$$

(iii) We use the finite, renormalized values for all of the superficially divergent nucleon loop diagrams illustrated in Fig. 12. These diagrams fall into six classes: (a) diagrams with external σ or π lines only, (b) diagrams with one axial-vector vertex and external meson lines, (c) diagrams with external photon lines only, (d) the axial-vector–photon–photon triangle diagram, (e) diagrams with external photon and meson lines, and (f) diagrams with an axial-vector vertex and external photon and meson lines. In Appendix A we give explicit renormalized expressions for the diagrams of types (a) and (b), and show that they satisfy the usual axial-vector Ward identities. The diagrams of type (c) (photon vacuum polarization loops) were considered in our discussion of spinor electrodynamics. The diagrams of types (d)–(f) are made finite and unique by calculating them in a gauge-invariant manner. As we have emphasized, the triangle diagram of type (d) does not satisfy the usual axial-vector Ward identity. We show in Appendix A that the axial-vector–photon–photon–meson box diagram [Fig. 12(f)] does satisfy the usual axial-vector Ward identity.

(iv) We take account of the counter term proportional to μ_0^2 in the following way. First, we omit *all* σ-meson tadpole diagrams (Fig. 9). (The recipe in Appendix A sets the basic nucleon loop tadpole equal to zero, but now tadpoles involving virtual meson integrations are to be dropped as well.) Second, when calculating pion self-energy diagrams $\Sigma^\pi(q^2)$ involving virtual meson integrations, a subtraction at $q=0$ should be performed to ensure that

$$\Sigma^\pi(0)=0. \tag{55}$$

[We will check explicitly below that the derivation of Eqs. (46)–(50) is valid in the cutoff theory.] This subtraction eliminates the formal quadratic divergence (which has become an actual logarithmic divergence in our cutoff theory) and leaves only formal logarithmic divergences, which are rendered finite by the cutoffs in

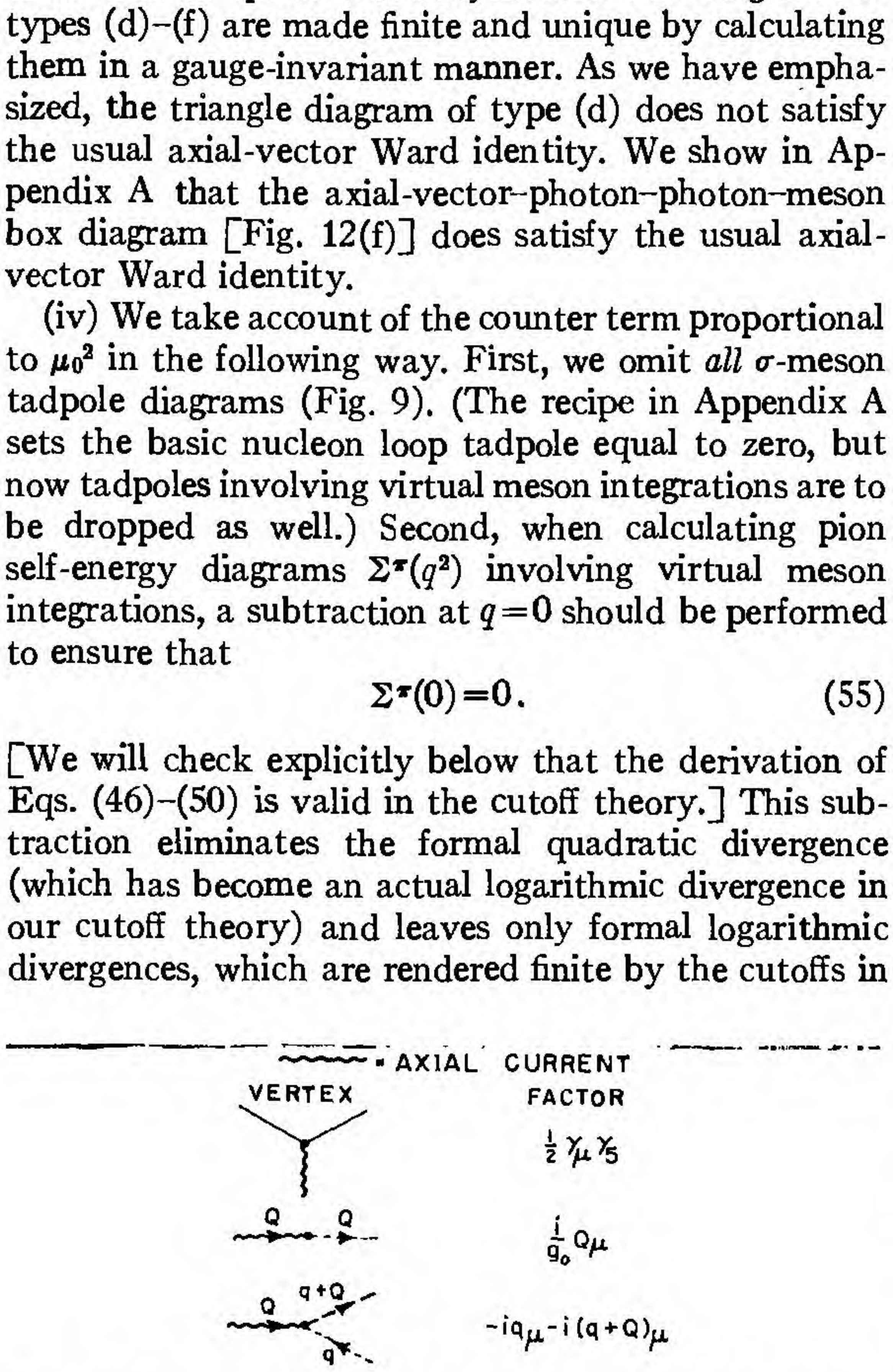


FIG. 11. Feynman rules for the σ model: axial-vector-current–nucleon and axial-vector-current–meson vertices.

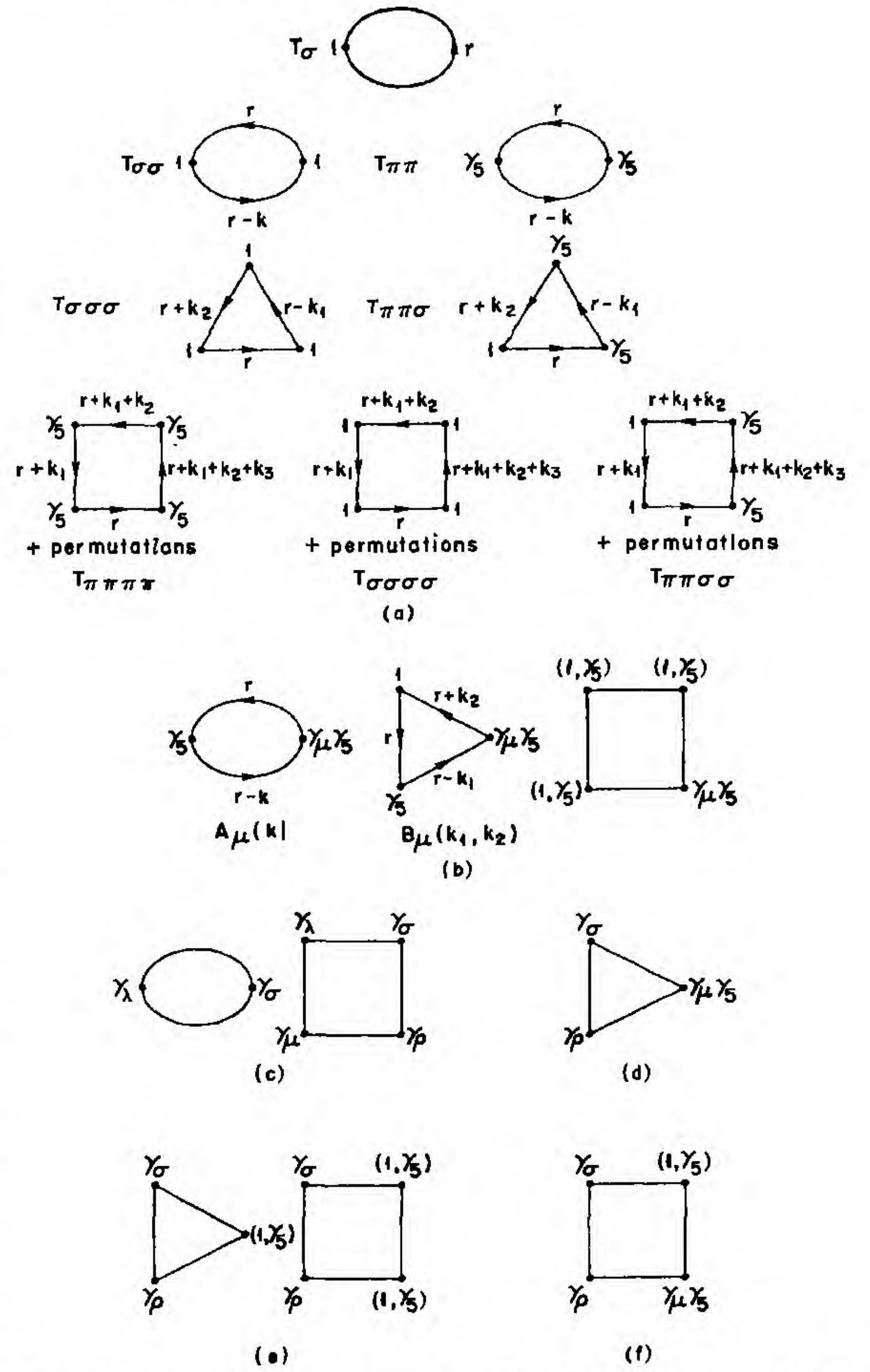


FIG. 12. Superficially divergent nucleon loop diagrams in the σ model. The six categories are described in the text immediately following Eq. (54).

the meson propagators. The σ self-energy is to be calculated from the pion self-energy by use of the equation

$$\Sigma^\sigma(q^2)=[\Sigma^\sigma(q^2)-\Sigma^\pi(0)]+\Sigma^\pi(0); \tag{56}$$

the quantity in square brackets is only formally logarithmically divergent, and hence finite in our cutoff theory. All other diagrams involving virtual meson integrations are automatically finite in the cutoff theory.

(v) There is a factor $\int d^4l/(2\pi)^4$ for each internal integration over loop variable l, a factor -1 for each fermion loop, a factor $\frac{1}{2}$ for each closed loop with one or two identical meson lines [Fig. 13(a)], and a factor $\frac{1}{6}$ for each closed loop with three identical meson lines [Fig. 13(b)].

(vi) We use the iterative renormalization procedure[10]

[10] We make no attempt to *prove* renormalizability of the σ model. Renormalizability of the σ model (with only the mesons present) has recently been discussed by B. W. Lee, Nucl. Phys B9, 649 (1969), and renormalization of the closely related φ^4 meson theory has been analyzed by T. T. Wu. Phys. Rev. **125**, 1436 (1962).

to fix the unrenormalized quantities e_0, g_0, G_0, λ_0, $m_0=G_0/g_0$, μ_1, and the wave-function renormalizations Z_2 (fermion wave-function renormalization), Z_3^π, Z_3^σ, and $Z_3^\gamma=(e/e_0)^{1/2}$. For finite Λ, all of these will be *finite* functions of Λ and of the renormalized quantities e, g, G, λ, and μ, with μ the physical pion mass. (Alternatively, we can take the independent physical quantities to be e, m, G, λ, and μ, with m the physical nucleon mass.) We include wave-function renormalization factors $Z_2^{1/2}$, $(Z_3^\pi)^{1/2}$, $(Z_3^\sigma)^{1/2}$, and $(Z_3^\gamma)^{1/2}$ for each fermion pion, σ, and photon external line.

As in the case of spinor electrodynamics, the cutoff rules in the σ model are compactly summarized by the statement that they are the Feynman rules for the regulated Lagrangian density[11]:

$$\begin{aligned}\mathfrak{L}^R(x)=&\bar\psi[i\partial-G_0(g_0^{-1}+\sigma^T+i\pi^T\gamma_5)]\psi+(D^{(2)}+\lambda_0)[4(\sigma^T)^2+4g_0\sigma^T((\sigma^T)^2+(\pi^T)^2)+g_0^2((\sigma^T)^2+(\pi^T)^2)^2]\\&+\tfrac{1}{2}E^{(2)}[(\partial\pi^T)^2+(\partial\sigma^T)^2]+\tfrac{1}{2}(F^{(2)}+\mu_0^2)[2g_0^{-1}\sigma^T+(\sigma^T)^2+(\pi^T)^2]+\tfrac{1}{2}[(\partial\pi)^2+(\partial\sigma)^2]-\tfrac{1}{2}\mu_1^2(\pi^2+\sigma^2)\\&-\tfrac{1}{2}[(\partial\pi^R)^2+(\partial\sigma^R)^2]+\tfrac{1}{2}\Lambda^2[(\pi^R)^2+(\sigma^R)^2]-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+\tfrac{1}{4}F_{\mu\nu}{}^RF^{R\mu\nu}-\tfrac{1}{2}\Lambda^2A_\mu{}^RA^{R\mu}-e_0\bar\psi\gamma_\mu\psi(A^\mu+A^{R\mu})\\&+C^{(2)}(F_{\mu\nu}+F_{\mu\nu}{}^R)(F^{\mu\nu}+F^{R\mu\nu}),\quad \pi^T=\pi+\pi^R,\quad \sigma^T=\sigma+\sigma^R.\end{aligned}\tag{57}$$

The axial-vector current, generated by the gauge transformation

$$\begin{aligned}\psi&\to(1+\tfrac{1}{2}i\gamma_5 v)\psi,\\ \pi&\to\pi-v(g_0^{-1}+\sigma),\\ g_0^{-1}+\sigma&\to g_0^{-1}+\sigma+v\pi,\\ \pi^R&\to\pi^R-v\sigma^R,\\ \sigma^R&\to\sigma^R+v\pi^R,\end{aligned}\tag{58}$$

is[12]

$$\begin{aligned}j_\mu{}^5=-\delta\mathfrak{L}^R/\delta(\partial^\mu v)=&\bar\psi\tfrac{1}{2}\gamma_\mu\gamma_5\psi+\sigma\partial_\mu\pi-\pi\partial_\mu\sigma+g_0^{-1}\partial_\mu\pi\\&-\sigma^R\partial_\mu\pi^R+\pi^R\partial_\mu\sigma^R+E^{(2)}\\&\times(\sigma^T\partial_\mu\pi^T-\pi^T\partial_\mu\sigma^T+g_0^{-1}\partial_\mu\pi^T).\end{aligned}\tag{59}$$

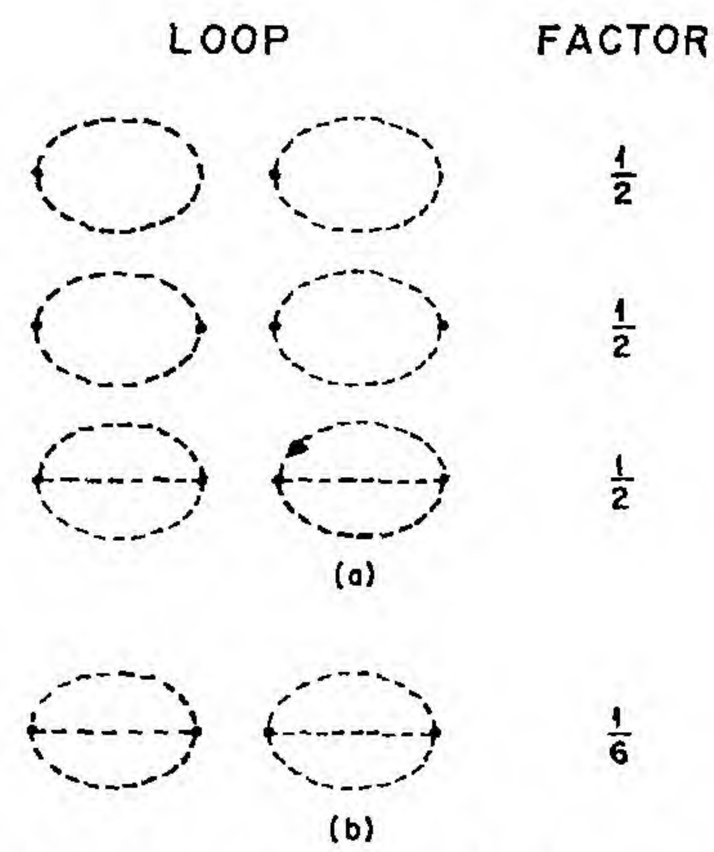


FIG. 13. Meson loop diagrams and corresponding Bose-symmetry factors.

The counter terms proportional to $D^{(2)}$, $E^{(2)}$, and $F^{(2)}$ perform the explicit subtractions in the loops illustrated in Figs. 12(a) and 12(b), just as $C^{(2)}$ provides the explicit subtraction in the basic vacuum polarization loop (see Appendix A for details).

We are now ready to calculate the divergence of the axial-vector current in our cutoff theory. One way to do this is to proceed diagrammatically, as we did in Eqs. (10)–(14) in the spinor electrodynamics case. However, because of the complexity of the σ model, this method will involve very complicated equations. Therefore, we will instead follow the second, more succinct, method used in spinor electrodynamics. We note first that calculation of the axial-vector divergence in the regulated σ model by naive use of the equations of motion gives

$$\partial^\mu j_\mu{}^5=-\delta\mathfrak{L}^R/\delta v=-(\mu_1^2/g_0)\pi.\tag{60}$$

Extra terms on the right-hand side of Eq. (60) can arise only from diagrams which are so singular that the Ward identities break down. However, since we have cut off the photon and meson propagators, all virtual boson integrations are strongly convergent and cannot lead to singularities which are not present when the boson integrations are omitted. Thus, breakdown of Eq. (60) can only be associated with the basic axial-vector loops shown in Figs. 12(b), 12(d), and 12(f). (All other axial-vector loops have enough vertices, and hence are convergent enough, to satisfy the normal axial-vector Ward identities.) By explicit calculation, we have found that of these diagrams, only the axial-vector–vector–vector triangle of Fig. 12(d) has an anomalous Ward identity, leading to the conclusion that, in the regulated σ model, the axial-vector-current divergence equation is

$$\begin{aligned}\partial^\mu j_\mu{}^5=-(\mu_1^2/g_0)\pi+\tfrac{1}{2}(\alpha_0/4\pi)(F^{\xi\sigma}+F^{R\xi\sigma})\\ \times(F^{\tau\rho}+F^{R\tau\rho})\epsilon_{\xi\sigma\tau\rho}.\end{aligned}\tag{61}$$

This completes our verification that Eq. (4) is exact to all orders of the strong and electromagnetic couplings in the σ model.

[11] Since $\langle\sigma\rangle_0=\langle\sigma^R\rangle_0=0$, the conditions $\langle\delta\mathfrak{L}/\delta\sigma\rangle_0=\langle\delta\mathfrak{L}/\delta\sigma^R\rangle_0=0$ are identical.

[12] The modified Feynman rules for meson propagators attached to the axial-vector vertex [item (ii) above] follow directly from Eq. (59). The pion propagator immediately following the axial-vector-current–pion vertex is unregulated because $g_0^{-1}\partial_\mu\pi$ appears in Eq. (59) without an accompanying $g_0^{-1}\partial_\mu\pi^R$ term. Similarly, the *product* of meson propagators following the axial-vector-current–pion–σ vertex is regulated as in Eq. (54) because the bilinear terms in Eq. (59) have the difference-of-products form $\sigma\partial_\mu\pi-\pi\partial_\mu\sigma-(\sigma^R\partial_\mu\pi^R-\pi^R\partial_\mu\sigma^R)$.

From Eq. (61) a number of consequences immediately follow.

(i) We can check the consistency of our cutoff Feynman rules by verifying that Eq. (55) is really valid in the regulated theory. As in Eq. (46), we define

$$\Delta_F^{\pi\prime}(q)=-i\int d^4x\, e^{iq\cdot x}\langle T(\pi(x)\pi(0))\rangle_0, \qquad (62)$$

and substitute Eq. (61) for $\pi(x)$. Using Eq. (59) for $j_\mu^5(x)$, and the canonical commutation relation

$$[\partial^0\pi(x)+E^{(2)}\partial^0\pi^T(x),\pi(0)]|_{x_0=0}=-i\delta^3(x), \qquad (63)$$

we still find the result $\Delta_F^{\pi\prime}(0)=-1/\mu_1^2$. The relation between $\Delta_F^{\pi\prime}(q)$ and the proper pion self-energy $\Sigma^\pi(q^2)$ in the cutoff theory is given by

$$\Delta_F^{\pi\prime}(q)=\frac{1}{q^2-\mu_1^2}+\frac{1}{q^2-\mu_1^2}\Sigma^\pi(q^2)\frac{1}{q^2-\mu_1^2}+\frac{1}{q^2-\mu_1^2}\Sigma^\pi(q^2)\left(\frac{1}{q^2-\mu_1^2}-\frac{1}{q^2-\Lambda^2}\right)\Sigma^\pi(q^2)\frac{1}{q^2-\mu_1^2}$$
$$+\frac{1}{q^2-\mu_1^2}\Sigma^\pi(q^2)\left(\frac{1}{q^2-\mu_1^2}-\frac{1}{q^2-\Lambda^2}\right)\Sigma^\pi(q^2)\left(\frac{1}{q^2-\mu_1^2}-\frac{1}{q^2-\Lambda^2}\right)\Sigma^\pi(q^2)\frac{1}{q^2-\mu_1^2}+\cdots$$
$$=\frac{1+\Sigma^\pi(q^2)(q^2-\Lambda^2)^{-1}}{q^2-\mu_1^2+\Sigma^\pi(q^2)(\mu_1^2-\Lambda^2)(q^2-\Lambda^2)^{-1}}, \qquad (64)$$

so that

$$\Delta_F^{\pi\prime}(0)=\frac{1-\Sigma^\pi(0)\Lambda^{-2}}{-\mu_1^2-\Sigma^\pi(0)(\mu_1^2-\Lambda^2)\Lambda^{-2}}, \qquad (65)$$

and therefore Eq. (49) still implies that $\Sigma^\pi(0)=0$.

(ii) In the absence of electromagnetism, Eq. (61) becomes the usual PCAC equation $\partial^\mu j_\mu^5=-(\mu_1^2/g_0)\pi$. From this equation, it is straightforward to prove[13] that the coupling-constant, mass, and wave-function renormalizations which make S-matrix elements in the σ model finite also make all matrix elements of j_μ^5 and of $(\mu_1^2/g_0)\pi$ finite (i.e., cutoff-independent as $\Lambda\to\infty$). In the presence of electromagnetism, the effect of the extra term in Eq. (61), as shown in I, is to induce an extra infinity in j_μ^5 which is *not* removed by the renormalizations which make the S matrix finite. However, just as we found that $m_0 j^5(x)$ is made finite by the fermion wave-function renormalization in spinor electrodynamics, we expect that, even in the presence of electromagnetism, $(\mu_1^2/g_0)\pi$ will be made finite by the pion wave-function renormalization in the σ model. That is, we expect

$$(\mu_1^2/g_0)(Z_3^\pi)^{1/2}=\text{finite}. \qquad (66)$$

Since the pion field is multiplicatively renormalizable,

$$\pi=(Z_3^\pi)^{1/2}\pi^{\text{renorm}}, \qquad (67)$$

to prove Eq. (66) we only need show that any particular nonvanishing matrix element of $(\mu_1^2/g_0)\pi$ is finite. The natural choice is the vacuum to two-photon matrix element $G^\pi(k_1\cdot k_2)$, defined by

$$\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|(-\mu_1^2/g_0)\pi|0\rangle$$
$$=(4k_{10}k_{20})^{-1/2}k_1{}^\xi k_2{}^\tau\epsilon_1{}^{*\sigma}\epsilon_2{}^{*\rho}\epsilon_{\xi\tau\sigma\rho}G^\pi(k_1\cdot k_2). \qquad (68)$$

[13] J. Bernstein, M. Gell-Mann, and L. Michel, Nuovo Cimento 16, 560 (1960); Preparata and Weisberger (Ref. 9).

Precisely the same arguments leading to Eq. (23) show that

$$G^\pi(0)=-\alpha/\pi, \qquad (69)$$

proving Eq. (66). We are now free to let the cutoff Λ approach infinity, defining a renormalized matrix element $\tilde{G}^\pi(k_1\cdot k_2)$,

$$\lim_{\Lambda\to\infty}G^\pi(k_1\cdot k_2)=\tilde{G}^\pi(k_1\cdot k_2), \qquad (70)$$

which satisfies the exact low-energy theorem

$$\tilde{G}^\pi(0)=-\alpha/\pi. \qquad (71)$$

In Sec. III, we will explicitly check Eq. (71) to second order in the strong meson-nucleon coupling constant G.

(iii) The low-energy theorem of Eq. (71) can be rewritten in a physically interesting form, as follows. We introduce the $\pi\to2\gamma$ decay amplitude $F^\pi(k_1\cdot k_2)$ and the pion weak decay amplitude f_π by writing

$$\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|(\Box^2+\mu^2)\pi^{\text{renorm}}|0\rangle$$
$$=(4k_{10}k_{20})^{-1/2}k_1{}^\xi k_2{}^\tau\epsilon_1{}^{*\sigma}\epsilon_2{}^{*\rho}\epsilon_{\xi\tau\sigma\rho}F^\pi(k_1\cdot k_2), \qquad (72)$$

and

$$\langle\pi(q)|j_\mu^5|0\rangle=(2q_0)^{-1/2}(-iq_\mu/\mu^2)f_\pi/\sqrt{2}. \qquad (73)$$

Comparing Eq. (72) with Eqs. (68) and (71), we find

$$\frac{-\mu_1^2}{g_0}\frac{(Z_3^\pi)^{1/2}}{\mu^2}F^\pi(0)=\frac{-\alpha}{\pi}, \qquad (74)$$

while taking the divergence of Eq. (73) and using Eq. (61) gives

$$f_\pi/\sqrt{2}=(-\mu_1^2/g_0)(Z_3^\pi)^{1/2}$$
$$+(\text{terms of higher order in }\alpha). \qquad (75)$$

Combining Eqs. (74) and (75) then gives the low-energy theorem relating the $\pi\to2\gamma$ and π weak decay ampli-

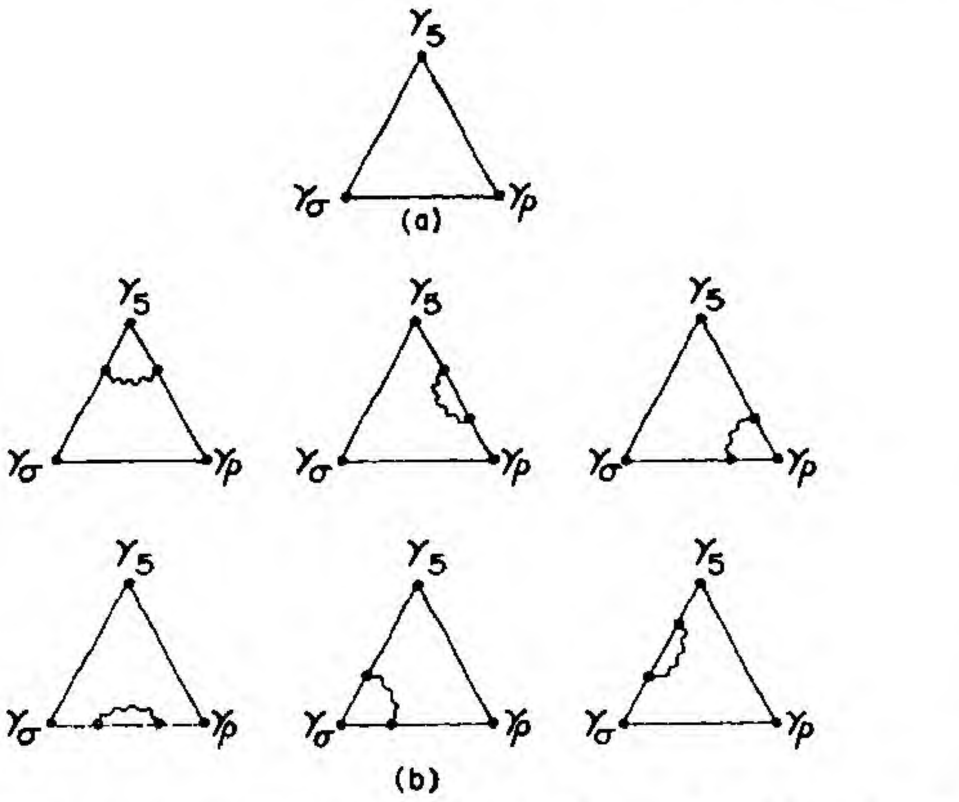


FIG. 14. (a) γ_5-γ_σ-γ_ρ skeleton triangle. (b) Second-order radiative corrections to the γ_5-γ_σ-γ_ρ triangle in spinor electrodynamics.

tudes,

$$F^\pi(0)=(-\alpha/\pi)(\sqrt{2}\mu^2/f_\pi)$$
$$+(\text{terms of higher order in } \alpha). \quad (76)$$

The fact that Eq. (61) is exact means that Eq. (76) is true to all orders in the strong interactions in the σ model. The experimental consequences of Eq. (76) are discussed in I.

III. SECOND-ORDER CALCULATION

We give in this section an explicit second-order calculation to check our contention that Eqs. (2) and (4) are exact. Rather than calculating corrections to both the axial-vector and pseudoscalar or pion vertices and checking Eqs. (2) and (4) directly, we will check these equations *indirectly* by verifying the low-energy theorems (30) and (71) which they imply. In the case of spinor electrodynamics, we will calculate the second-order radiative corrections to the matrix element $\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|2im_0\bar{\psi}\gamma_5\psi|0\rangle$, which arise from the six diagrams shown in Fig. 14(b). These diagrams [plus mass counter terms appearing in the basic γ_5-γ_σ-γ_ρ triangle of Fig. 14(a)] make a contribution to $\tilde{G}(0)$ of order α^2, which must, in fact, be zero for Eq. (30) to be correct. The vanishing of the α^2 term is clearly a test of the absence of a term proportional to $\alpha\alpha_0 F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$ in Eq. (2).

Similarly, in the σ model we will calculate radiative corrections to $\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|(-\mu_1^2/g_0)\pi|0\rangle$. It is convenient to rewrite this matrix element by substituting for $\mu_1^2\pi$ the pion equation of motion obtained from Eq. (31),

$$\mu_1^2\pi=-\Box^2\pi-iG_0\bar{\psi}\gamma_5\psi$$
$$+\lambda_0[8g_0\sigma\pi+4g_0^2\pi(\sigma^2+\pi^2)]+\mu_0^2\pi. \quad (77)$$

The matrix element of $\Box^2\pi$ makes a contribution to $\tilde{G}^\pi(k_1\cdot k_2)$ of order $k_1\cdot k_2$, and thus can be neglected at $k_1\cdot k_2=0$. If we work to second order in G^2 but to zeroth order in λ, so that the physical pion and σ masses remain equal, the meson-meson scattering terms in Eq. (77) can also be dropped. Finally, let us recall that the effect of the counter term proportional to μ_0^2 is to produce a subtraction in the pion proper self-energy, giving $\Sigma^\pi(0)=0$. In particular, this means that the counter term $\mu_0^2\pi$ in Eq. (77) combines with the nucleon bubble diagram involving $-iG_0\bar{\psi}\gamma_5\psi$, shown in Fig. 15, to give a contribution to $\tilde{G}^\pi(k_1\cdot k_2)$ proportional to $\Sigma^\pi(2k_1\cdot k_2)$, which vanishes at $k_1\cdot k_2=0$. Recalling that $G_0/g_0=m_0$, we may summarize the findings of this paragraph by the statement that to check the low-energy theorem of Eq. (71) to order G^2, we need only calculate the matrix element $\langle\gamma(k_1,\epsilon_1)\gamma(k_2,\epsilon_2)|im_0\bar{\psi}\gamma_5\psi|0\rangle$, omitting the bubble diagram of Fig. 15. The twelve diagrams which contribute have the form of those in Fig. 14(b), with the virtual photon line replaced by a virtual pion or a virtual σ line. These diagrams, plus mass counter terms in the basic triangle of Fig. 14(a), make a contribution to $\tilde{G}^\pi(0)$ of order $G^2\alpha$, which must vanish for Eq. (71) to be valid [thereby verifying the absence of a term proportional to $G^2\alpha_0 F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$ in Eq. (14)]. Thus we see that the second-order spinor electrodynamics and σ-model calculations will appear very similar.

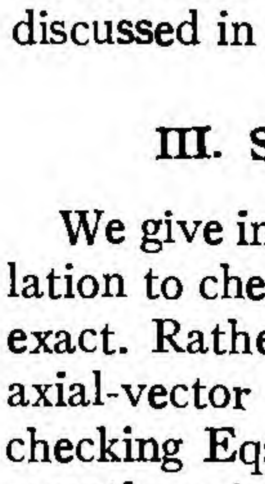


FIG. 15. Nucleon bubble diagram, which in the σ model appears in the second-order radiative corrections to the γ_5-γ_σ-γ_ρ triangle.

The calculations of the second-order radiative corrections to the triangle diagram will proceed in the following way. First, we calculate the renormalized quantities $\tilde{\Gamma}_\mu(p,p')$, $\tilde{\Gamma}^5(p,p')$, and $\tilde{S}_F'(p)$ in spinor electrodynamics and the σ model, and substitute them into the γ_5-γ_σ-γ_ρ skeleton triangle of Fig. 14(a). The constants $\tilde{G}(0)$ and $\tilde{G}^\pi(0)$ are determined by extracting the first nonvanishing terms of a Taylor series expansion of the amplitudes in the photon momenta k_1 and k_2. The Ward identities and integrations by parts are used to show that the self-energy corrections are exactly canceled by the unexpanded vertex corrections. The terms where an external momentum has been expanded from a vertex correction function are evaluated in two ways: by direct calculation and by using a further integration by parts. Using either method, the sum of these terms is found to vanish. Therefore, the second-order radiative corrections to $\tilde{G}(0)$ and $\tilde{G}^\pi(0)$ are zero.

Let $\tilde{\Gamma}_\mu^{(2)}(p,p')$, $\tilde{\Gamma}_5^{(2)}(p,p')$, and $\tilde{S}_F'^{(2)}(p)$ denote the renormalized second-order vector vertex, pseudoscalar vertex, and fermion propagator in either spinor electrodynamics or the σ model, as defined by Eqs. (25) and (28). (Explicit expressions for these quantities are given in Appendix B.) The matrix element which we want

is proportional to

$$\mathfrak{M}_{\sigma\rho}=\int d^4r\,\mathrm{Tr}[\tilde{\Gamma}_5{}^{(2)}(r-k_2,r+k_1)\tilde{S}_F{}'^{(2)}(r+k_1)\times\tilde{\Gamma}_\sigma{}^{(2)}(r+k_1,r)\tilde{S}_F{}'^{(2)}(r)\tilde{\Gamma}_\rho{}^{(2)}(r,r-k_2)\times\tilde{S}_F{}'^{(2)}(r-k_2)]. \quad (78)$$

Since we are actually only interested in studying the part of Eq. (78) coming from the second-order radiative corrections, let us substitute

$$\begin{aligned}\tilde{\Gamma}_\mu{}^{(2)}(p,p')&=\gamma_\mu+\Lambda_\mu(p,p'),\\ \tilde{\Gamma}_5{}^{(2)}(p,p')&=\gamma_5+\Lambda_5(p,p'),\\ \tilde{S}_F{}'^{(2)}(p)&=[\not{p}-m-\Sigma(\not{p})]^{-1}\end{aligned}\quad (79)$$

into Eq. (78) and isolate the second-order part. This gives

$$\begin{aligned}\mathfrak{M}_{\sigma\rho}{}^{(2)}=\int d^4r\,\mathrm{Tr}[&\Lambda_5(r-k_2,r+k_1)(\not{r}+\not{k}_1-m)^{-1}\gamma_\sigma(\not{r}-m)^{-1}\gamma_\rho(\not{r}-\not{k}_2-m)^{-1} &(80a)\\ &+\gamma_5(\not{r}+\not{k}_1-m)^{-1}\Sigma(\not{r}+\not{k}_1)(\not{r}+\not{k}_1-m)^{-1}\gamma_\sigma(\not{r}-m)^{-1}\gamma_\rho(\not{r}-\not{k}_2-m)^{-1} &(80b)\\ &+\gamma_5(\not{r}+\not{k}_1-m)^{-1}\Lambda_\sigma(r+k_1,r)(\not{r}-m)^{-1}\gamma_\rho(\not{r}-\not{k}_2-m)^{-1} &(80c)\\ &+\gamma_5(\not{r}+\not{k}_1-m)^{-1}\gamma_\sigma(\not{r}-m)^{-1}\Sigma(\not{r})(\not{r}-m)^{-1}\gamma_\rho(\not{r}-\not{k}_2-m)^{-1} &(80d)\\ &+\gamma_5(\not{r}+\not{k}_1-m)^{-1}\gamma_\sigma(\not{r}-m)^{-1}\Lambda_\rho(r,r-k_2)(\not{r}-\not{k}_2-m)^{-1} &(80e)\\ &+\gamma_5(\not{r}+\not{k}_1-m)^{-1}\gamma_\sigma(\not{r}-m)^{-1}\gamma_\rho(\not{r}-\not{k}_2-m)^{-1}\Sigma(\not{r}-\not{k}_2)(\not{r}-\not{k}_2-m)^{-1}], &(80f)\end{aligned}$$

with the six terms in Eq. (80) corresponding, of course, to the six diagrams in Fig. 14(b). To evaluate Eq. (80), we use the fact that although the integral over r is apparently linearly divergent, the linearly divergent parts of terms (80a)–(80f) vanish separately when the trace is taken.[14] This means that we can simplify the form of Eq. (80) by making separate translations of the integration variable in each of the pieces (80a)–(80f), as follows:

(a) $r\to r+k_2$, (d) $r\to r$,
(b) $r\to r-k_1$, (e) $r\to r$,
(c) $r\to r-k_1$, (f) $r\to r+k_2$.

After making these translations, wherever Σ appears it has argument r, and the vertex parts Λ_5, Λ_σ, and Λ_ρ have r as the first argument. Next, we Taylor-expand with respect to k_1 and k_2, keeping only the leading term of order k_1k_2 (because of the γ_5, the terms of order 1, k_1, k_2, $k_1{}^2$, and $k_2{}^2$ vanish identically). Defining the vertex derivatives $\Lambda_{5,\xi}(r)$ and $\Lambda_{\sigma,\xi}(r)$ by

$$\begin{aligned}\Lambda_5(r,r+a)&=\Lambda_5(r,r)+a^\xi\Lambda_{5,\xi}(r)+O(a^2),\\ \Lambda_\sigma(r,r+a)&=\Lambda_\sigma(r,r)+a^\xi\Lambda_{\sigma,\xi}(r)+O(a^2),\end{aligned}\quad (81)$$

we find

$$\mathfrak{M}_{\sigma\rho}{}^{(2)}=\mathfrak{M}_{A\sigma\rho}{}^{(2)}+\mathfrak{M}_{B\sigma\rho}{}^{(2)}+(\text{terms of higher order in momenta}), \quad (82)$$

with [we abbreviate $s\equiv(\not{r}-m)^{-1}$]

$$\begin{aligned}\mathfrak{M}_{A\sigma\rho}{}^{(2)}&\equiv k_1{}^\xi k_2{}^\tau\mathfrak{M}_{A\xi\tau\sigma\rho}{}^{(2)}\\ &=\int d^4r\,\mathrm{Tr}[\Lambda_5(r,r)s(-\not{k}_1)s\gamma_\sigma s(-\not{k}_2)s\gamma_\rho s+\gamma_5 s\Sigma(\not{r})s\gamma_\sigma s\not{k}_1 s\gamma_\rho s\not{k}_2 s+\gamma_5 s\Lambda_\sigma(r,r)s\not{k}_1 s\gamma_\rho s\not{k}_2 s\\ &\quad+\gamma_5 s(-\not{k}_1)s\gamma_\sigma s\Sigma(\not{r})s\gamma_\rho s\not{k}_2 s+\gamma_5 s(-\not{k}_1)s\gamma_\sigma s\Lambda_\rho(r,r)s\not{k}_2 s+\gamma_5 s(-\not{k}_1)s\gamma_\sigma s(-\not{k}_2)s\gamma_\rho s\Sigma(\not{r})s], \quad (83)\end{aligned}$$

and with

$$\begin{aligned}\mathfrak{M}_{B\sigma\rho}{}^{(2)}&\equiv k_1{}^\xi k_2{}^\tau\mathfrak{M}_{B\xi\tau\sigma\rho}{}^{(2)}\\ &=\int d^4r\,\mathrm{Tr}[k_1{}^\xi\Lambda_{5,\xi}(r)s\gamma_\sigma s(-\not{k}_2)s\gamma_\rho s+\gamma_5 s(-k_1{}^\xi)\Lambda_{\sigma,\xi}(r)s\gamma_\rho s\not{k}_2 s+\gamma_5 s(-\not{k}_1)s\gamma_\sigma s(-k_2{}^\tau)\Lambda_{\rho,\tau}(r)s]. \quad (84)\end{aligned}$$

The tensors $\mathfrak{M}_{A\xi\tau\sigma\rho}{}^{(2)}$ and $\mathfrak{M}_{B\xi\tau\sigma\rho}{}^{(2)}$ must both have the structure

$$\begin{aligned}\mathfrak{M}_{A\xi\tau\sigma\rho}{}^{(2)}&=\epsilon_{\xi\tau\sigma\rho}\mathfrak{M}_A{}^{(2)},\\ \mathfrak{M}_{B\xi\tau\sigma\rho}{}^{(2)}&=\epsilon_{\xi\tau\sigma\rho}\mathfrak{M}_B{}^{(2)},\end{aligned}\quad (85)$$

and therefore are completely antisymmetric in their indices. Clearly, each *individual term* in the sums in Eqs. (83) and (84) will also have the form of Eq. (85),

[14] Simple γ-matrix counting shows that trace of each of the terms (80a)–(80f) is proportional to the fermion mass m, and is therefore, on dimensional grounds, at most logarithmically divergent. Because of the over-all factor γ_5, the logarithmic divergences vanish as well, so that the integrals of terms (80a)–(80f) actually converge.

once the integration over r is performed. This fact greatly simplifies the following calculation.

To evaluate $\mathfrak{M}_{A\xi\tau\sigma\rho}^{(2)}$, we eliminate the vertex parts from Eq. (83) by using the Ward identities

$$\begin{aligned} \Lambda_5(r,r) &= (1/2m)[\gamma_5\Sigma(r)+\Sigma(r)\gamma_5], \\ \Lambda_\lambda(r,r) &= -\partial_\lambda\Sigma(r), \\ -\partial_\lambda(r-m)^{-1} &= (r-m)^{-1}\gamma_\lambda(r-m)^{-1}. \end{aligned} \tag{86}$$

Integration by parts in the terms involving $\partial_\lambda\Sigma(r)$ shifts the derivatives to the free propagators $(r-m)^{-1}$. The resultant amplitudes all involve a trace containing γ matrices, r, and $\Sigma(r)$. By anticommuting r through the γ matrices and using the total antisymmetry of $\mathfrak{M}_{A\xi\tau\sigma\rho}^{(2)}$ in its tensor indices, we find

$$\begin{aligned} \mathfrak{M}_{A\xi\tau\sigma\rho}^{(2)} = \int d^4r \, \{&(r^2-m^2)^{-3} \\ &\times \mathrm{Tr}[\tfrac{1}{2}(\gamma_5\Sigma(r)+\Sigma(r)\gamma_5)\gamma_\xi\gamma_\sigma\gamma_\tau\gamma_\rho] \\ &+(r^2-m^2)^{-4}\,\mathrm{Tr}[(r+m)\Sigma(r)(r+m)\gamma_5\gamma_\xi\gamma_\sigma\gamma_\tau\gamma_\rho] \\ &-(r^2-m^2)^{-4}8r_\sigma\,\mathrm{Tr}[\Sigma(r)(r+m)\gamma_5\gamma_\xi\gamma_\tau\gamma_\rho]\}. \end{aligned} \tag{87}$$

On substitution of the general expression $\Sigma(r)=A(r^2)+rB(r^2)$ and use of symmetric integration in r, we find that the right-hand side of Eq. (87) vanishes, implying $\mathfrak{M}_{A\xi\tau\sigma\rho}^{(2)}=0$.

The evaluation of $\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}$ in Eq. (84) will be done using two different methods, each giving zero. The first method involves a direct calculation of the integrals. We recall that each term of Eq. (84) is totally antisymmetric in the tensor indices, once the integration over r is performed. Using this total antisymmetry and reversing the order of the γ matrix products in the third term in Eq. (84) yields

$$\begin{aligned} \mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)} = \int d^4r\, \mathrm{Tr}\{&\Lambda_{5,\xi}(r)s\gamma_\sigma s(-\gamma_\tau)s\gamma_\rho s \\ &+\gamma_5 s[-\Lambda_{\sigma,\xi}(r)+\Lambda_{\sigma,\xi}(r)^R]s\gamma_\rho s\gamma_\tau s\}, \end{aligned} \tag{88}$$

where $\Lambda_{\sigma,\xi}^{(2)}(r)^R$ is obtained from $\Lambda_{\sigma,\xi}^{(2)}(r)$ by reversing the order of all γ-matrix products. From Appendix B, we find the expressions

$$\begin{aligned} \Lambda_{5,\xi}(r) &= \gamma_5\gamma_\xi E_1(r^2)+\gamma_5 r_\xi E_2(r^2), \\ \Lambda_{\sigma,\xi}(r) &= \gamma_\sigma r_\xi D_1(r^2)+\gamma_\xi\gamma_\sigma r D_2(r^2)+r\gamma_\sigma\gamma_\xi D_3(r^2) \\ &\quad +g_{\sigma\xi}D_4(r^2)+r_\sigma r_\xi D_5(r^2)+r_\sigma r_\xi r D_6(r^2), \\ \Lambda_{\sigma,\xi}(r)^R &= \gamma_\sigma r_\xi D_1(r^2)+r\gamma_\sigma\gamma_\xi D_2(r^2)+\gamma_\xi\gamma_\sigma r D_3(r^2) \\ &\quad +g_{\sigma\xi}D_4(r^2)+r_\sigma r_\xi D_5(r^2)+r_\sigma r_\xi r D_6(r^2), \end{aligned} \tag{89}$$

where E_1, E_2, $D_1,\cdots$, are simple integrals over Feynman parameters. Substituting Eq. (89) into Eq. (88) and evaluating the trace gives

$$\mathfrak{M}_B^{(2)} = -4i\int d^4r\,\{m^2E_1(r^2) -mr^2[D_2(r^2)-D_3(r^2)]\}(r^2-m^2)^{-3}. \tag{90}$$

From Appendix B we find

$$\begin{aligned} E_1(r^2) &= \frac{1}{16\pi^2}\begin{Bmatrix} e^2 \\ -G^2 \end{Bmatrix} \times\int_0^1 zdz\frac{-2m}{-r^2z(1-z)+zm^2+(1-z)\mu^2}, \\ D_2(r^2)-D_3(r^2) &= \frac{1}{16\pi^2}\begin{Bmatrix} e^2 \\ -G^2 \end{Bmatrix} \times\int_0^1 zdz\frac{2(1-z)}{-r^2z(1-z)+zm^2+(1-z)\mu^2}, \end{aligned} \tag{91}$$

where the upper (lower) entry in { } refers to spinor electrodynamics (the σ model), and where μ^2 is the virtual photon, pion, or σ mass. Inserting Eq. (91) into Eq. (90), we find that the integral in Eq. (90) is proportional to $I(\mu^2/m^2)$, with

$$I(a)\equiv\int_0^1 zdz\int_0^\infty \frac{udu}{(u+1)^3}\frac{1-u(1-z)}{uz(1-z)+z+(1-z)a}. \tag{92}$$

We show in Appendix C that this integral is identically zero, giving $\mathfrak{M}_B^{(2)}=0$.

The second method used to evaluate $\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}$ involves the use of an integration by parts. Since the derivative on the vertex function removes the effect of the renormalization constants, the three terms in Eq. (84) may be written diagrammatically as shown in Fig. 16(a). In the first term in Eq. (84), we use Eq. (86) to replace $(r-m)^{-1}\gamma_\rho(r-m)^{-1}$ by $-\partial_\rho(r-m)^{-1}$, and then integrate by parts, using the total antisymmetry of the amplitude to drop the terms in which the derivative acts on the propagators adjacent to k_1 and γ_σ. This operation has the effect of replacing the left-hand diagram in Fig. 16(a) by the diagram in Fig. 16(b). The expression for $\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}$ becomes

$$\begin{aligned} \mathfrak{M}_{B\sigma\rho}^{(2)} &= k_1{}^\xi k_2{}^\tau\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)} \\ &= \int d^4r\,\mathrm{Tr}[\gamma_5 s(-k_1)sk_2{}^\tau\Lambda_{\sigma,\tau}(r)s(-\gamma_\rho)s \\ &\quad +\gamma_5 s(-k_1{}^\xi)\Lambda_{\sigma,\xi}(r)s\gamma_\rho sk_2 s \\ &\quad +\gamma_5 s(-k_1)s\gamma_\sigma s(-k_2{}^\tau)\Lambda_{\rho,\tau}(r)s]. \end{aligned} \tag{93}$$

Each term in the sum of Eq. (95) involves a trace containing γ matrices, r, and the function $\Lambda_{\sigma,\xi}(r)$. By anticommuting r through the γ matrices and by using

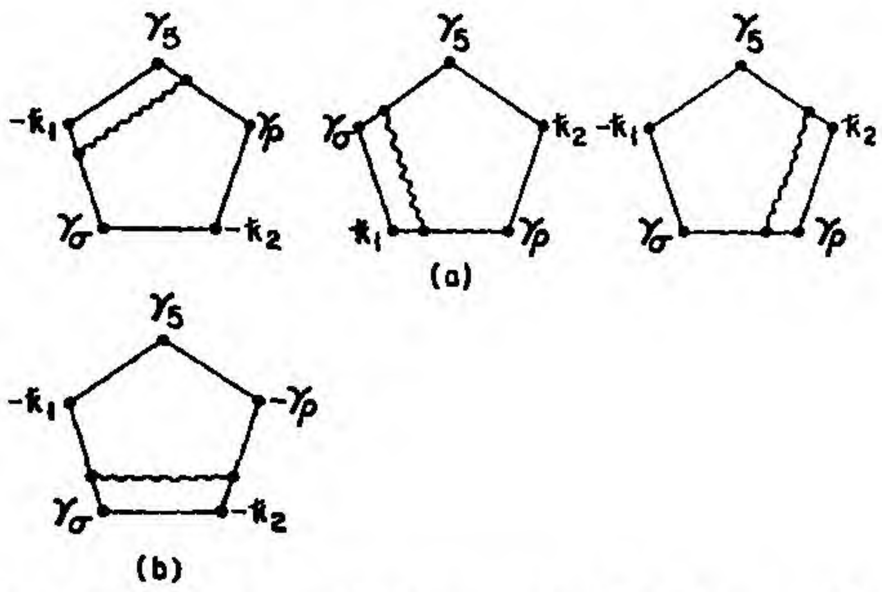


FIG. 16. (a) Diagrammatic representation of $\mathfrak{M}_{B\sigma\rho}^{(2)}$ in spinor electrodynamics. (b) Diagram obtained from the left-hand loop in (a) by integration by parts.

the antisymmetry of $\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}$ in its indices, one finds

$$\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}=\int d^4r\,(r^2-m^2)^{-2}\,\mathrm{Tr}\{\gamma_5\Lambda_{\sigma,\rho}(r)\tfrac{1}{2}[\gamma_\xi,\gamma_\tau]\}\,. \tag{94}$$

(The terms coming from the anticommutators exactly cancel because of the antisymmetry.) Substituting Eq. (89) into Eq. (94) immediately gives $\mathfrak{M}_{B\xi\tau\sigma\rho}^{(2)}=0$. It is not actually necessary to have the detailed form of Eq. (89) to see that Eq. (94) vanishes. Referring to Fig. 17(a), we see that in spinor electrodynamics $\Lambda_{\sigma,\rho}$ has the form

$$\Lambda_{\sigma,\rho}=\gamma_\alpha\Lambda_{\sigma,\rho}'(r)\gamma^\alpha\,, \tag{95}$$

and when Eq. (95) is substituted into Eq. (94), the sum over virtual photon polarization states α cancels to zero. Similarly, from Fig. 17(b) we see that in the σ model $\Lambda_{\sigma,\rho}$ has the property[15]

$$\Lambda_{\sigma,\rho}(r)=i\gamma_5\Lambda_{\sigma,\rho}(r)i\gamma_5\,, \tag{96}$$

which again implies that Eq. (94) vanishes. In this case, the virtual pion term is exactly canceled by the virtual σ term.

IV. SUMMARY AND DISCUSSION

We summarize our results and briefly compare them with the recent findings of Jackiw and Johnson.[2] We have considered two models, spinor electrodynamics and a truncated version of the σ model. In each model, we have studied a *particular* axial-vector current: in spinor electrodynamics, the usual axial-vector current of Eq. (1), and in the σ model, the Polkinghorne[16] axial-vector current of Eq. (3), which, in the absence of electromagnetism, obeys the PCAC condition. By introducing cutoffs in the boson propagators we have shown that, in the presence of electromagnetism, the divergences of our axial-vector currents are modified in a simple well-defined way, to *all* orders of perturbation theory. The modification consists of the addition of a simple numerical mutliple of $(\alpha_0/4\pi)F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$ to the naive axial-vector divergence. (The naive divergence is the one obtained by formal manipulation of equations of motion when subtleties arising from the singularity of local-field products are neglected.) From the anomalous divergence equations we obtained simple low-energy theorems for the vacuum-to-2γ matrix element of the naive divergence. Although these theorems were derived with the cutoff Λ finite, we argued that in both models, even in the presence of electromagnetism, the naive divergence is a finite (cutoff-independent) operator as $\Lambda\to\infty$.[17] This allowed us to pass freely to the limit $\Lambda\to\infty$, obtaining a low-energy theorem for the renormalized naive divergence operator. This low-energy theorem was checked explicitly to second order in radiative corrections in a calculation using only renormalized (cutoff-independent) quantities, verifying our contention that no subtleties were involved in the $\Lambda\to\infty$ limit.

Thus, in our calculation, use of the cutoff has been an artifice, and the cutoff does not appear in the physics. As is made clear by the discussion of Eqs. (66)–(70), this important feature can be traced directly back to the following property of the two axial-vector currents which we have studied: *The naive divergences of the two axial-vector currents, as well as the axial-vector currents themselves, are multiplicatively renormalizable.* We conjecture that in any renormalizable theory field with an axial-vector current satisfying this property, arguments analogous to those of this paper can be carried through.

With these comments in mind, let us examine the conclusions of Jackiw and Johnson. Jackiw and Johnson treat the electromagnetic field as an external (nonquantized) variable, but allow quantized strong interactions of the spinor particles, so their calculation applies, for example, to the σ model. Rather than considering the Polkinghorne current of Eq. (3), Jackiw and Johnson take as the axial-vector current the fermion part $\bar{\psi}\gamma_\mu\gamma_5\psi$ alone. They find that the effect of the strong interactions on the anomalous divergence term is ambiguous, and depends on precisely how a cutoff is introduced. It is easy to see that (even in the absence of electromagnetism) the current $\bar{\psi}\gamma_\mu\gamma_5\psi$ is *not* made finite by the usual renormalizations which make

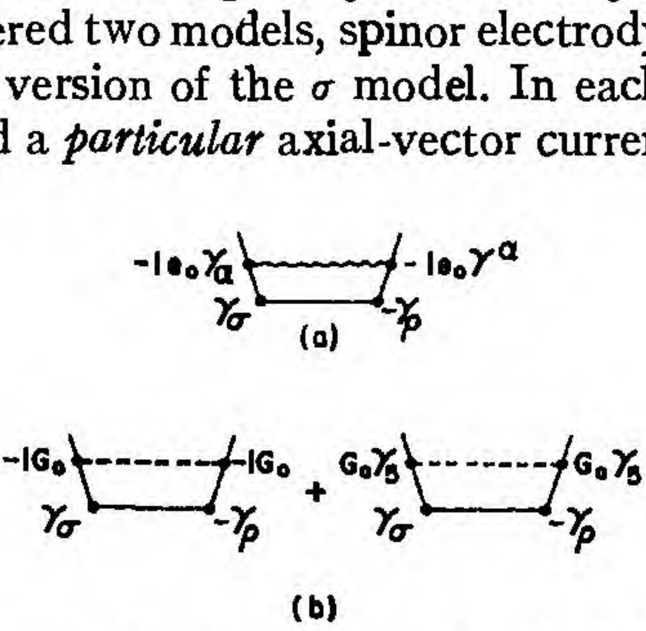


FIG. 17. Diagrammatic representation of $\Lambda_{\sigma,\rho}$ in (a) spinor electrodynamics and (b) the σ model.

[15] Comparing Eq. (96) with Eq. (89), we see that in the σ model, we must have $D_4=D_5=0$. This can be verified from the explicit formulas of Appendix B.

[16] J. C. Polkinghorne, Nuovo Cimento 8, 179 (1958); 8, 781 (1958).

[17] That is, multiplication by the usual external-line wave-function renormalization factors makes Feynman amplitudes of the naive divergence finite.

the S matrix in the σ model finite, and, by a reversal of the Preparata-Weisberger argument,[18] this means that the naive divergence of this current is *not* multiplicatively renormalizable. In other words, the axial-vector current considered by Jackiw and Johnson and its naive divergence are not well-defined objects in the usual renormalized perturbation theory; hence the ambiguous results which these authors have obtained are not too surprising.

The presence of two different axial-vector currents in the σ model poses, however, the following question: Which current should we take as the prototype for the *physical* axial-vector current? The answer is that there are two arguments in favor of using the full Polkinghorne current, rather than its fermion part alone, as the current to which the semileptonic weak interactions couple: (i) We want the physical axial-vector current to satisfy the PCAC hypothesis. Although PCAC does not require that the divergence of the axial-vector current be a canonical pion field (as is the case for the Polkinghorne current), it does require that it at least be a *smooth* interpolating field for the pion. However, a non-multiplicatively renormalizable operator, such as the divergence of the current $\bar{\psi}\gamma_\mu\gamma_5\psi$, will *not* be a smooth operator and therefore is not a pion interpolating field suitable for PCAC arguments. (ii) When charged fields and charged currents are added to the model, we want the axial-vector currents to satisfy the Gell-Mann current-algebra hypothesis. As Gell-Mann and Lévy[8] have shown, the Polkinghorne axial-vector current is the one which obeys the current algebra. The fermion part alone does not satisfy the current algebra.

To summarize, then, in the σ model (and also in the neutral-vector-meson model, which behaves like spinor electrodynamics), the current which is a prototype for the physical axial-vector current has a simple anomalous divergence term in the presence of electromagnetism. Other axial-vector currents can be defined which do not have simple anomalous divergence behavior, but these currents are not good prototypes for the physical current.

ACKNOWLEDGMENTS

We wish to thank N. Christ, R. F. Dashen, M. L. Goldberger, and S. B. Treiman for helpful conversations, and R. Jackiw and K. Johnson for a stimulating controversy which led to the writing of this paper. One of us (W.A.B.) wishes to thank Dr. Carl Kaysen for the hospitality of the Institute for Advanced Study.

APPENDIX A

In the first part of this Appendix, we give explicit renormalized expressions for the diagrams depicted in Figs. 12(a) and 12(b), and we show that they satisfy the usual axial-vector Ward identities. In the second part, we demonstrate that the axial-vector–photon–photon–meson box diagram of Fig. 12(f) satisfies the usual axial-vector Ward identity.[19]

A. Meson and Axial-Vector-Current–Meson Loops

In order to give unambiguous values to the divergent loops which we encounter, we define the symmetric, cutoff integral symbol

$$\int_{[C,M^2]} . \tag{A1}$$

Equation (A1) is a shorthand for the following sequence of operations: (i) We do the usual Dyson rotation to the Euclidean region; (ii) we integrate symmetrically over the angle variables around the center $r=C$; and (iii) we integrate the Euclidean magnitude squared $x=-r^2$ up to an upper limit of M^2. Using this symbol, we define *unrenormalized* meson and axial-vector-current–meson loops as follows[20]:

Meson Loops

$$T_\sigma=-G_0\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r-m_0}\right),$$

$$T_{\sigma\sigma}(k)=-i\Sigma_\sigma(k),\quad T_{\pi\pi}(k)=-i\Sigma_\pi(k),$$

$$\Sigma_\sigma(k)=-iG_0^2\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r-m_0}\frac{1}{r-k-m_0}\right),$$

$$\Sigma_\pi(k)=iG_0^2\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\gamma_5\frac{1}{r-m_0}\gamma_5\frac{1}{r-k-m_0}\right),$$

[18] Preparata and Weisberger (Ref. 9), Appendix C.

[19] For a derivation of the Ward identities satisfied by the general spinor loop coupling to scalar, pseudoscalar, vector, and axial-vector external sources, with full SU_3 structure, see W. Bardeen (to be published). The results of Appendix A are a special case of the general problem discussed by Bardeen.

[20] In this Appendix, $\Sigma_\pi(k)$ and $\Sigma_\sigma(k)$ denote, respectively, the pion and σ proper self-energies, which were denoted by $\Sigma^\pi(k^2)$ and $\Sigma^\sigma(k^2)$ in the text.

$$T_{\pi\pi\sigma}(k_1,k_2)=2G_0{}^3\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\gamma_5\frac{1}{r-k_1-m_0}\gamma_5\frac{1}{r+k_2-m_0}\frac{1}{r-m_0}\right),$$

$$T_{\sigma\sigma\sigma}(k_1,k_2)=-2G_0{}^3\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r-k_1-m_0}\frac{1}{r+k_2-m_0}\frac{1}{r-m_0}\right),$$

$$T_{\pi\pi\pi\pi}(k_1,k_2,k_3)=-G_0{}^4\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\gamma_5\frac{1}{r+k_1+k_2+k_3-m_0}\gamma_5\frac{1}{r+k_1+k_2-m_0}\right.$$
$$\left.\times\gamma_5\frac{1}{r+k_1-m_0}\gamma_5\frac{1}{r-m_0}+\text{five permutations}\right),$$

$$T_{\sigma\sigma\sigma\sigma}(k_1,k_2,k_3)=-G_0{}^4\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r+k_1+k_2+k_3-m_0}\frac{1}{r+k_1+k_2-m_0}\right.$$
$$\left.\times\frac{1}{r+k_1-m_0}\frac{1}{r-m_0}+\text{five permutations}\right),$$

$$T_{\pi\pi\sigma\sigma}(k_1,k_2,k_3)=G_0{}^4\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\gamma_5\frac{1}{r+k_1+k_2+k_3-m_0}\gamma_5\frac{1}{r+k_1+k_2-m_0}\right.$$
$$\left.\times\frac{1}{r+k_1-m_0}\frac{1}{r-m_0}+\text{five permutations}\right).\quad\text{(A2)}$$

Axial-Vector-Current-Meson Loops

$$A_\mu(k)=-G_0\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\tfrac{1}{2}\gamma_\mu\gamma_5\frac{1}{r-m_0}\gamma_5\frac{1}{r-k-m_0}\right)+\frac{iG_0m_0}{(4\pi)^2}k_\mu$$
$$=-G_0\int_{[0,M^2e^{1/2}]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\tfrac{1}{2}\gamma_\mu\gamma_5\frac{1}{r-m_0}\gamma_5\frac{1}{r-k-m_0}\right),\quad\text{(A3)}$$

$$B_\mu(k_1,k_2)=2G_0{}^2\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\gamma_5\frac{1}{r-k_1-m_0}\tfrac{1}{2}\gamma_\mu\gamma_5\frac{1}{r+k_2-m_0}\frac{1}{r-m_0}\right).$$

The loops T_σ, $T_{\sigma\sigma}$, $T_{\pi\pi}$, A_μ, and B_μ *are* at least linearly divergent, and so specification of the center for symmetric averaging is essential. The three-meson and four-meson loops, on the other hand, are not linearly divergent, and so the origin of integration in these loops may be freely translated. (Terms which vanish as $M\to\infty$ will be picked up from the translation, but may be ignored. Similarly, the two different expressions which we have given for A_μ are not precisely equal, but differ by terms which vanish as $M\to\infty$.)

We have chosen identical upper limits M^2 for all loop integrals except A_μ, where the upper limit has been taken as $M^2e^{1/2}$ (e is the base of natural logarithms). These choices of upper limit guarantee that the unrenormalized loops satisfy the usual axial-vector Ward identities. For example, in the case of A_μ, a straightforward calculation shows that

$$k^\mu(-G_0)\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\tfrac{1}{2}\gamma_\mu\gamma_5\frac{1}{r-m_0}\gamma_5\frac{1}{r-k-m_0}\right)$$
$$=-\frac{m_0}{G_0}i[\Sigma_\pi(k)-\Sigma_\pi(0)]+\tfrac{1}{2}G_0I(k),\quad\text{(A4)}$$

with

$$I(k)=\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r-k-m_0}-\frac{1}{r-m_0}\right)$$
$$=\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4}\mathrm{Tr}\left(\frac{1}{r-k-m_0}k\frac{1}{r-m_0}k\frac{1}{r-m_0}\right)$$
$$=\frac{-2im_0k^2}{(4\pi)^2}+O(M^{-2}).\quad\text{(A5)}$$

Therefore, $A_\mu(k)$, as defined in Eq. (A3), satisfies the usual axial-vector Ward identity

$$k^\mu A_\mu(k) = -(m_0/G_0)i[\Sigma_\pi(k) - \Sigma_\pi(0)]. \quad \text{(A6)}$$

Similarly, the axial-vector triangle $B_\mu(k_1,k_2)$ satisfies the Ward identity

$$-(k_1+k_2)^\mu B_\mu(k_1,k_2) = (m_0/G_0)T_{\pi\pi\sigma} + i\Sigma_\sigma(k_2) - i\Sigma_\pi(k_1). \quad \text{(A7)}$$

The axial-vector-current–three-meson box diagram of Fig. 12(b) is superficially logarithmically divergent, but, because $\mathrm{Tr}\{\gamma_\mu\gamma_5 r(1,\gamma_5)r(1,\gamma_5)r(1,\gamma_5)r\}=0$, this diagram actually converges, which is why we have not included it in the list of unrenormalized axial-vector-current loops in Eq. (A3). Introducing a cutoff at $x=M^2$ (which changes this diagram only by terms which vanish as $M\to\infty$) and then taking the divergence yields a linear combination of three- and four-meson loop diagrams, all with cutoff at $x=M^2$ as in Eq. (A2) Some of these loops may occur with the loop integration variable translated by a finite amount with respect to the standard forms in Eq. (A2), but as we have pointed out, this does not matter because none of the three- or four-meson loops is linearly divergent. We conclude, then, that the axial-vector-current–three-meson box diagram and the meson loop diagrams of Eq. (A2) satisfy the usual axial-vector Ward identity. Identical reasoning shows that the axial-vector-current–four-meson pentagon, which is superficially convergent, is related by the usual Ward identity to a linear combination of the meson box diagrams of Eq. (A2) and to the convergent meson pentagon diagram. Note that because the axial-vector-current box and pentagon diagrams are *finite*, their Ward identities will necessarily involve linear combinations of the meson triangle and box diagrams in which the logarithmic divergences exactly cancel.

Having defined the unrenormalized loop diagrams and shown that they satisfy the correct Ward identities, we next construct the renormalized loops and show that they, too, satisfy the proper Ward identities. The renormalized meson scattering and axial-vector-current loops are obtained from the unrenormalized loops by adding appropriate matrix elements of $i\mathcal{L}^{\mathrm{counter}}$ and $j_\mu{}^{5\,\mathrm{counter}}$ with [see Eqs. (57) and (59)]

$$\mathcal{L}^{\mathrm{counter}} = D^{(2)}[4\sigma^2 + 4(G_0/m_0)\sigma(\sigma^2+\pi^2) + (G_0/m_0^2)(\sigma^2+\pi^2)^2] + \tfrac{1}{2}E^{(2)}[(\partial\pi)^2+(\partial\sigma)^2] + \tfrac{1}{2}F^{(2)}[(2m_0/G_0)\sigma+\sigma^2+\pi^2], \quad \text{(A8)}$$

$$j_\mu{}^{5\,\mathrm{counter}} = E^{(2)}[\sigma\partial_\mu\pi - \pi\partial_\mu\sigma + (m_0/G_0)\partial_\mu\pi].$$

The subtractions $D^{(2)}$, $E^{(2)}$, and $F^{(2)}$ are determined from the second-order π and σ self-energy diagrams to be

$$D^{(2)} = \frac{m_0^2G_0^2}{(4\pi)^2}\ln\left(\frac{M^2}{m_0^2}\right) + \tilde{D}^{(2)},$$

$$E^{(2)} = \frac{-2G_0^2}{(4\pi)^2}\ln\left(\frac{M^2}{m_0^2}\right) + \tilde{E}^{(2)}, \quad \text{(A9)}$$

$$F^{(2)} = \Sigma_\pi(0) = iG_0^2\int_{[0,M^2]}\frac{d^4r}{(2\pi)^4} \times \mathrm{Tr}\left(\gamma_5\frac{1}{r-m_0}\gamma_5\frac{1}{r-m_0}\right).$$

The finite constants $\tilde{D}^{(2)}$ and $\tilde{E}^{(2)}$ are adjusted to give the physical pion mass and the meson-meson scattering constant the specified values μ^2 and λ. The renormalized loops, denoted by a tilde, are given by

$$\begin{aligned}
\tilde{T}_\sigma &= T_\sigma + i(m_0/G_0)F^{(2)} = 0,\\
\tilde{\Sigma}_\pi(k) &= \Sigma_\pi(k) - F^{(2)} - k^2E^{(2)},\\
\tilde{\Sigma}_\sigma(k) &= \Sigma_\sigma(k) - F^{(2)} - 8D^{(2)} - k^2E^{(2)},\\
\tilde{T}_{\pi\pi\sigma} &= T_{\pi\pi\sigma} + 8i(G_0/m_0)D^{(2)},\\
\tilde{T}_{\sigma\sigma\sigma} &= T_{\sigma\sigma\sigma} + 24i(G_0/m_0)D^{(2)},\\
\tilde{T}_{\pi\pi\pi\pi} &= T_{\pi\pi\pi\pi} + 24i(G_0^2/m_0^2)D^{(2)},\\
\tilde{T}_{\sigma\sigma\sigma\sigma} &= T_{\sigma\sigma\sigma\sigma} + 24i(G_0^2/m_0^2)D^{(2)},\\
\tilde{T}_{\pi\pi\sigma\sigma} &= T_{\pi\pi\sigma\sigma} + 8i(G_0^2/m_0^2)D^{(2)},\\
\tilde{A}_\mu(k) &= A_\mu(k) + i(m_0/G_0)E^{(2)}k_\mu,\\
\tilde{B}_\mu(k_1,k_2) &= B_\mu(k_1,k_2) + iE^{(2)}(k_2-k_1)_\mu.
\end{aligned} \quad \text{(A10)}$$

It is straightforward to verify that all of the tilde quantities approach finite limits as $M\to\infty$, showing, as required by chiral invariance, that the subtraction constants determined from the second-order loops make the triangle and box diagrams finite as well.

From Eqs. (A6), (A7), and (A10) we find that the renormalized loops $\tilde{A}_\mu$ and $\tilde{B}_\mu$ satisfy the desired Ward identities

$$\begin{aligned}
k^\mu\tilde{A}_\mu(k) &= -i(m_0/G_0)\tilde{\Sigma}_\pi(k),\\
-(k_1+k_2)^\mu\tilde{B}_\mu(k_1,k_2) &= (m_0/G_0)\tilde{T}_{\pi\pi\sigma} + i\tilde{\Sigma}_\sigma(k_2) - i\tilde{\Sigma}_\pi(k_1).
\end{aligned} \quad \text{(A11)}$$

Next, we recall that the axial-vector-current box and pentagon diagrams and the unrenormalized meson-scattering triangle and box diagrams are related by the usual Ward identities. This implies that the same Ward identities are satisfied by the axial-vector box and pentagon and the *renormalized* meson-scattering diagrams. The reason is that the counter terms which subtract the divergences in the meson loops necessarily occur in each Ward identity in the same linear combination as the logarithmic divergences, and therefore cancel among themselves in the Ward identity in the same manner as the logarithmic divergences do. This com-

pletes the proof that the renormalized basic loop diagrams satisfy the normal axial-vector Ward identities.

B. Axial-Vector-Current–Photon–Photon–Meson Loop [Figure 12(f)]

The Ward identity for the axial-vector-current–photon–photon–meson loop of Fig. 12(f) relates it to a linear combination of the diagrams shown in Fig. 12(e), plus a possible anomalous term. To see that no anomalous term is actually present, we note that: (i) Gauge invariance makes the diagrams of Figs. 12(e) and 12(f) finite, so no renormalizations are needed to make the various terms in the Ward identity well defined; (ii) a possible anomalous term must be gauge-invariant with respect to both photon indices, must be odd in m_0, and must have the dimensions of a mass, since all of the other terms in the Ward identity have these properties; (iii) a possible anomalous term can have *no singularities* in internal masses or external momentum variables, since taking an absorptive part eliminates the superficially logarithmically or linearly divergent loop integrals, and therefore the absorptive parts satisfy the usual Ward identities. According to (ii), a possible anomalous term must have the form

$$\text{anomalous term} \propto m_0 F_1 F_2 / g(m_0, \text{external momenta}), \tag{A12}$$

where F_1 and F_2 are the two photon field strength tensors and where g has the dimensions of $(\text{mass})^2$. However, because of the division by g, Eq. (A12) necessarily has singularities, and therefore (iii) forces the anomalous term to be zero.

APPENDIX B

We state here the renormalized second-order vector vertex, pseudoscalar vertex, and fermion propagator used in the calculation of Sec. III.

$$\tilde{\Gamma}_\lambda^{(2)}(p,p')=\gamma_\lambda+\frac{c^2}{16\pi^2}\int_0^1 zdz\int_0^1 dy\left\{2\gamma_\lambda \ln\left[\frac{z^2m^2+(1-z)\mu^2}{D}\right]-\frac{N_\lambda}{D}-\frac{2m^2\gamma_\lambda P_1}{z^2m^2+(1-z)\mu^2}\right\},$$

$$\tilde{\Gamma}_5^{(2)}(p,p')=\gamma_5+\frac{c^2}{16\pi^2}\int_0^1 zdz\int_0^1 dy\left\{8\gamma_5 f \ln\left[\frac{z^2m^2+(1-z)\mu^2}{D}\right]+\frac{\gamma_5 N}{D}+\frac{4m^2\gamma_5 P_2}{z^2m^2+(1-z)\mu^2}\right\},$$

$$\tilde{S}_F'^{(2)}(p)=[\not p-m-\Sigma(\not p)]^{-1}, \tag{B1}$$

$$\Sigma(\not p)=\frac{c^2}{16\pi^2}\int_0^1 zdz\left\{2g_1\ln\left[\frac{z^2m^2+(1-z)\mu^2}{-p^2z(1-z)+zm^2+(1-z)\mu^2}\right]\right.$$
$$\left.+g_2\frac{m^2-p^2(1-z)^2}{-p^2z(1-z)+zm^2+(1-z)\mu^2}+\frac{2m^2\not p P_1+4m^3P_2}{z^2m^2+(1-z)\mu^2}\right\},$$

$$D=(y^2z^2-yz)p^2+[(1-y)^2z^2-(1-y)z]p'^2+2y(1-y)z^2p\cdot p'+zm^2+(1-z)\mu^2.$$

The quantities c, N_λ, N, P_1, P_2, f, g_1, and g_2 are defined as follows:

Spinor Electrodynamics

$$\begin{aligned}
&c=e,\\
&N_\lambda=-2m^2\gamma_\lambda-2[(1-z+yz)\not p'-yz\not p]\\
&\qquad\times\gamma_\lambda[(1-yz)\not p-(1-y)z\not p']+4m[(1-2yz)p_\lambda\\
&\qquad\qquad+(1-2z+2yz)p_\lambda'],\\
&N=4m^2-4[(1-yz)\not p-(1-y)z\not p']\cdot\\
&\qquad\times[(1-z+yz)\not p'-yz\not p]+2m(\not p-\not p'),\\
&P_1=z^2+2z-2,\quad P_2=1-2z,\\
&f=1,\\
&g_1=4m-\not p,\quad g_2=4m-2\not p.
\end{aligned} \tag{B2}$$

σ Model

$$\begin{aligned}
&c=G,\\
&N_\lambda=-2m^2\gamma_\lambda-2[(1-yz)\not p-(1-y)z\not p']\\
&\qquad\times\gamma_\lambda[(1-z+yz)\not p'-yz\not p],\\
&N=-2m(\not p-\not p'),\\
&P_1=z^2-2z+2,\quad P_2=-(1-z)^2,\\
&f=0,\\
&g_1=-\not p,\quad g_2=-2\not p.
\end{aligned} \tag{B3}$$

APPENDIX C

We show that the integral

$$I(a)\equiv\int_0^1 zdz\int_0^\infty \frac{udu}{(u+1)^3}\frac{1-u(1-z)}{uz(1-z)+z+(1-z)a} \tag{C1}$$

is identically zero. We begin by observing that $I(a)$ is analytic in the a plane, apart from a cut along the real axis from 0 to $-\infty$. The discontinuity across this cut at $a=-A$ is proportional to

$$\rho(A)\equiv\int_0^1\frac{zdz}{1-z}\int_0^\infty\frac{udu}{(u+1)^3}\times[1-u(1-z)]\delta(uz+z/(1-z)-A)$$
$$=\int_0^{A/(A+1)}dz\frac{z(1-z)[A-(A+1)z][(2+A)z-A]}{[-z^2+A(1-z)]^3}$$
$$=-\frac{1}{2A}\int_0^{A/(A+1)}dz\frac{d}{dz}\left\{\frac{z^2[A-(A+1)z]^2}{[-z^2+A(1-z)]^2}\right\}=0, \tag{C2}$$

which means that $I(a)$ is an entire function. Furthermore, for $\mathrm{Re}\,a\geq 0$,

$$|I(a)|\leq\int_0^1 dz\int_0^\infty\frac{udu}{(u+1)^3}\left|\frac{z+uz(1-z)}{uz(1-z)+z+(1-z)a}\right|$$
$$\leq\int_0^1 dz\int_0^\infty\frac{udu}{(u+1)^3}=\frac{1}{2} \tag{C3}$$

and

$$I(0)=2\int_0^\infty\frac{udu}{(u+1)^3}\int_0^1\frac{dz}{u(1-z)+1}-\int_0^\infty\frac{udu}{(u+1)^3}=\frac{1}{2}-\frac{1}{2}=0. \tag{C4}$$

Since Feynman integrals like Eq. (C1) never lead to functions of exponential type, Eqs. (C1)-(C4) show that $I(a)\equiv 0$.

PHYSICAL REVIEW D VOLUME 4, NUMBER 11 1 DECEMBER 1971

Comments and Addenda

The Comments and Addenda section is for short communications which are not of such urgency as to justify publication in Physical Review Letters *and are not appropriate for regular Articles. It includes only the following types of communications: (1) comments on papers previously published in* The Physical Review *or* Physical Review Letters; *(2) addenda to papers previously published in* The Physical Review *or* Physical Review Letters, *in which the additional information can be presented without the need for writing a complete article. Manuscripts intended for this section should be accompanied by a brief abstract for information-retrieval purposes. Accepted manuscripts will follow the same publication schedule as articles in this journal, and galleys will be sent to authors.*

Low-Energy Theorem for $\gamma+\gamma\rightarrow\pi+\pi+\pi$

Stephen L. Adler,* Benjamin W. Lee,† and S. B. Treiman‡
National Accelerator Laboratory, Batavia, Illinois 60510

and

A. Zee
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 10 September 1971; revised manuscript received 1 October 1971)

We use the hypothesis of the partially conserved axial-vector current (PCAC) to show that the matrix elements for $\gamma+\gamma\rightarrow\pi^0+\pi^0+\pi^0$ and $\gamma+\gamma\rightarrow\pi^0+\pi^++\pi^-$ vanish in the soft-π^0 limit. This, combined with photon gauge invariance, implies low-energy theorems relating these matrix elements to the matrix elements for $\gamma+\gamma\rightarrow\pi^0$ and $\gamma\rightarrow\pi^0+\pi^++\pi^-$. Since the magnitude of the former is determined by the π^0 lifetime, while the ratio of the latter to the former is determined in a model-independent way by isospin and low-energy-theorem arguments, a model-independent prediction for the $\gamma+\gamma\rightarrow\pi+\pi+\pi$ amplitude can be given. Our results agree with those of Aviv, Hari Dass, and Sawyer in the neutral case, but not in the charged case. We give a diagrammatic and effective-Lagrangian interpretation of our formulas which explains the discrepancy.

The reaction $\gamma+\gamma\rightarrow\pi+\pi+\pi$ is of interest, both because it may be observable in electron-positron colliding-beam experiments,[1] and because it is relevant to theoretical unitarity calculations[2] of a lower bound on the decay rate of $K_L^0\rightarrow\mu^+\mu^-$ In recent papers, Aviv, Hari Dass, and Sawyer[3] and Yao[4] have applied effective-Lagrangian methods to calculate the matrix elements for the neutral and charged cases of $\gamma+\gamma\rightarrow\pi+\pi+\pi$. The fact that Refs. 3 and 4 are in disagreement has prompted us to repeat the calculation by standard current-algebra–PCAC methods.[5] Our results agree with Ref. 3 (but not with Ref. 4) in the neutral case $\gamma+\gamma\rightarrow\pi^0+\pi^0+\pi^0$, and disagree with *both* Refs. 3 and 4 in the more interesting charged case $\gamma+\gamma\rightarrow\pi^0+\pi^++\pi^-$. After briefly discussing our method and results, we explain the reasons for our disagreement with the earlier calculations.

We begin with the simple, but powerful observation that the matrix elements

$$\mathfrak{M}^{0+-}\equiv\mathfrak{M}(\gamma(k_1)+\gamma(k_2)\rightarrow\pi^0(q_0)+\pi^+(q_+)+\pi^-(q_-))$$

and

$$\mathfrak{M}^{000}\equiv\mathfrak{M}(\gamma(k_1)+\gamma(k_2)\rightarrow\pi^0(q_0)+\pi^0(q_0')+\pi^0(q_0''))$$

vanish in the single-soft-π^0 limit $q_0\rightarrow 0$, with the remaining two pions held on the mass shell. To see this, we follow the standard PCAC procedure[6] of writing the reduction formula describing $\mathfrak{M}^{0+-}$ or $\mathfrak{M}^{000}$ with the π^0 off the mass shell, and then replacing the π^0 field by the divergence of the axial-vector current $(M_\pi^2 f)^{-1}\,\partial_\lambda\mathfrak{F}_3^{5\lambda}$. [The normalization constant f is given by $f\approx f_\pi/(\sqrt{2}M_\pi^2)\approx 0.68\,M_\pi$, with f_π the charged-pion decay amplitude.] Because the corresponding axial charge F_3^5 commutes with the electromagnetic current, *no* equal-time commutator terms are picked up when the derivative ∂_λ is brought outside the T product in the reduction formula. Integration by parts then makes the derivative act on the π^0 wave function, producing a factor $q_{0\lambda}$. Thus both $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$ are proportional to q_0, and since they contain no pole terms which become singular as $q_0\rightarrow 0$, they vanish in this limit. Note that this argument is unaltered by the divergence anomaly[7] in $\partial_\lambda\mathfrak{F}_3^{5\lambda}$, since when $\mathfrak{F}_3^{5\lambda}$ is the only axial-vector current

present, its divergence anomaly vanishes when the associated four-momentum q_0 vanishes.[8,9]

In addition to the soft-π^0 limit which we have just derived, we know that $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$ must be gauge-invariant. That is, they are bilinear forms in ϵ_1 and ϵ_2 (the polarization vectors of the two photons) and vanish when either ϵ_1 is replaced by k_1 or ϵ_2 is replaced by k_2. We can now invoke the standard lore of current-algebra low-energy theorems,[5] which tells us that since we know three independent pieces of information about the low-energy behavior of $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$ (the $q_0 \to 0$ limit, gauge invariance for photon 1, and gauge invariance for photon 2), we can determine $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$ from their pion-pole diagrams up to an error of order $O(q_0k_1k_2)$ at least.[10] In particular, the terms in $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$ *quadratic* in the momenta k_1, k_2, q_0, $q_+(q_0')$, and $q_-(q_0'')$ are completely determined. The relevant pion-pole diagrams are illustrated in Fig. 1. The pion-pion scattering amplitudes which appear are evaluated from the current-algebra expression[11,12]

$$\mathfrak{M}(\pi^a \to \pi^b(q_b)+\pi^c(q_c)+\pi^d(q_d)) = if^{-2}\{\delta_{bc}\delta_{ad}[(q_b+q_c)^2 - M_\pi^2] + \delta_{bd}\delta_{ac}[(q_b+q_d)^2 - M_\pi^2] + \delta_{cd}\delta_{ab}[(q_c+q_d)^2 - M_\pi^2]$$
$$- x[(q_b+q_c)^2+(q_b+q_d)^2+(q_c+q_d)^2 - 3M_\pi^2](\delta_{bc}\delta_{ad}+\delta_{bd}\delta_{ac}+\delta_{cd}\delta_{ab})\}, \quad (1a)$$

where x is a parameter proportional to the isotensor component of the "σ term" and is related to the $I=0$ pion-pion S-wave scattering length a_0 by

$$a_0 = (7/32\pi)f^{-2}M_\pi(1-\tfrac{5}{7}x). \quad (1b)$$

The $\gamma+\gamma\to\pi^0$ and $\gamma\to\pi^0+\pi^++\pi^-$ amplitudes are expressed in terms of coupling constants F^π and $F^{3\pi}$ defined by

$$\mathfrak{M}(\gamma(k_1)+\gamma(k_2)\to\pi^0) = ik_1^\alpha k_2^\beta \epsilon_1^\gamma \epsilon_2^\delta \epsilon_{\alpha\beta\gamma\delta}F^\pi,$$
$$\mathfrak{M}(\gamma(k_1)\to\pi^0+\pi^+(q_+)+\pi^-(q_-)) = ik_1^\alpha \epsilon_1^\beta q_+^\gamma q_-^\delta \epsilon_{\alpha\beta\gamma\delta}F^{3\pi}. \quad (2)$$

The coupling constant F^π is related to the π^0 lifetime by[13]

$$\tau_{\pi^0}^{-1} = (M_\pi^3/64\pi)(F^\pi)^2. \quad (3)$$

Comparison with experiment gives $|F_\pi| = (\alpha/\pi)(0.66\pm0.08M_\pi)^{-1}$, with α the fine-structure constant. While the coupling constant $F^{3\pi}$ has not been measured, both the theory of PCAC anomalies[14] and model-independent isospin and low-energy-theorem arguments (see below) predict

$$eF^{3\pi} = f^{-2}F^\pi, \quad e=(4\pi\alpha)^{1/2}. \quad (4)$$

Combining Eqs. (1) and (2) with the appropriate propagators to form the pion-pole diagrams, and adding the unique second-degree polynomial which guarantees gauge invariance and vanishing of the matrix elements as $q_0\to 0$, we get the following predictions for $\mathfrak{M}^{0+-}$ and $\mathfrak{M}^{000}$:

$$\mathfrak{M}^{000} = (1-3x)\overline{\mathfrak{M}}(q_0, q_0', q_0''),$$
$$\overline{\mathfrak{M}}(q_0,q_0',q_0'') = if^{-2}F^\pi k_1^\alpha k_2^\beta \epsilon_1^\gamma \epsilon_2^\delta \epsilon_{\alpha\beta\gamma\delta}\left(1 - \frac{(q_0+q_0')^2+(q_0+q_0'')^2+(q_0'+q_0'')^2-3M_\pi^2}{(q_0+q_0'+q_0'')^2-M_\pi^2}\right)$$
$$= if^{-2}F^\pi k_1^\alpha k_2^\beta \epsilon_1^\gamma \epsilon_2^\delta \epsilon_{\alpha\beta\gamma\delta}\left(\frac{-M_\pi^2}{(k_1+k_2)^2-M_\pi^2}\right) \quad \text{(when three final pions are on mass shell)}, \quad (5a)$$

$$\mathfrak{M}^{0+-} = if^{-2}F^\pi k_1^\alpha k_2^\beta \epsilon_1^\gamma \epsilon_2^\delta \epsilon_{\alpha\beta\gamma\delta}\left(1-\frac{(q_++q_-)^2-M_\pi^2}{(q_0+q_++q_-)^2-M_\pi^2}\right) - x\overline{\mathfrak{M}}(q_0,q_+,q_-)$$
$$-ieF^{3\pi}\epsilon_1^\gamma\epsilon_2^\delta\left[\left(\frac{(2q_+-k_2)_\delta}{k_2^2-2q_+\cdot k_2}k_1^\alpha(q_+-k_2)^\sigma q_-^\tau - \frac{(2q_--k_2)_\delta}{k_2^2-2q_-\cdot k_2}k_1^\alpha q_+^\sigma(q_--k_2)^\tau\right)\epsilon_{\alpha\gamma\sigma\tau}\right.$$
$$\left.+(k_1-k_2,\gamma-\delta)+(k_1-k_2)^\alpha q_0^\tau\epsilon_{\alpha\gamma\delta\tau}\right]. \quad (5b)$$

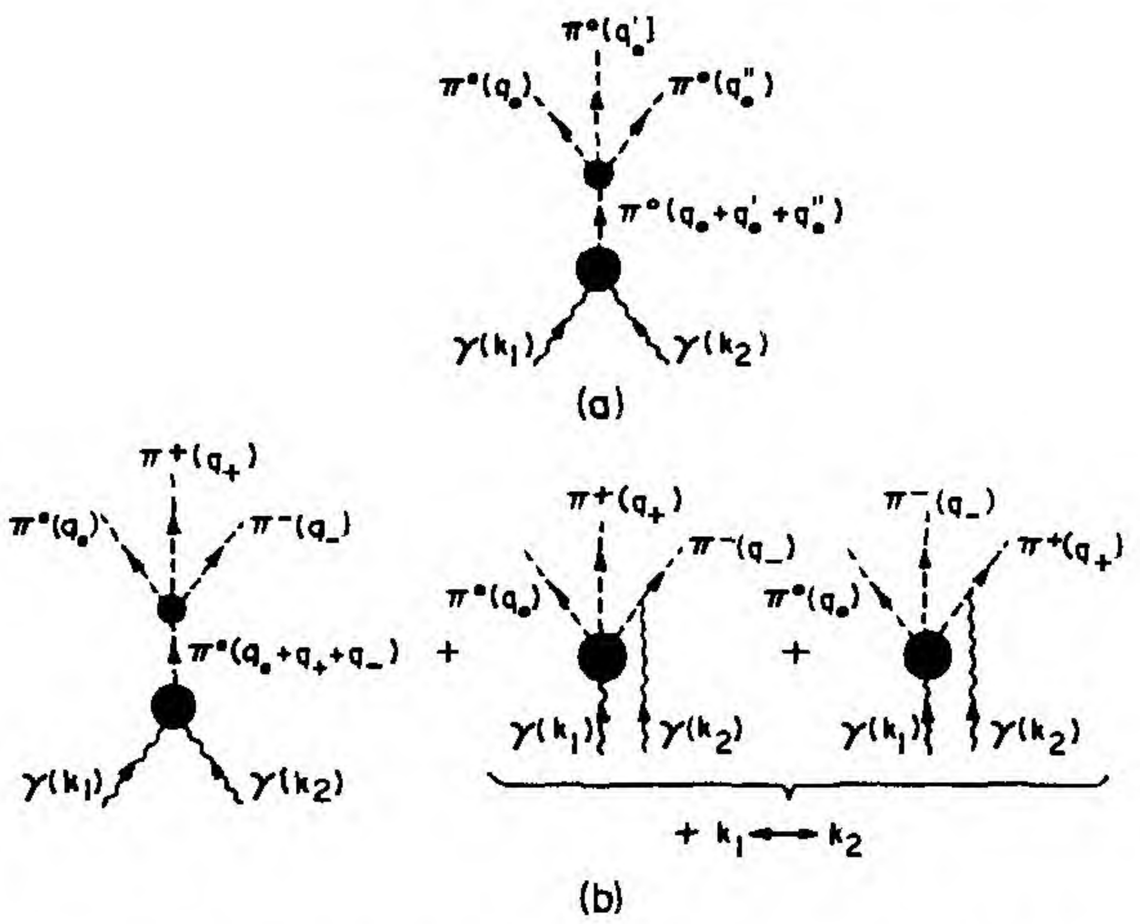


FIG. 1. Pion-pole diagrams for (a) the neutral and (b) the charged cases.

These equations are our basic results.[15]

Our expression for $\mathfrak{M}^{000}$ in Eq. (5a) agrees with that given by Aviv *et al.* We disagree with the result for $\mathfrak{M}^{000}$ quoted by Yao, who has (through an apparent algebraic error) replaced $-M_\pi^2$ in Eq. (5a) by $-4M_\pi^2$. In the case of strictly massless pions, our on-shell result for $\mathfrak{M}^{000}$ is the simple statement that the terms in the matrix element quadratic in the external momenta vanish.[16] This result can be immediately generalized to the reaction $\gamma+\gamma\to n\pi^0$, as follows: The PCAC argument given above tells us that in the limit when any one π^0 has zero four-momentum, with the other $n-1$ π^0's on the mass shell, the matrix element $\mathfrak{M}(\gamma+\gamma\to n\pi^0)$ must vanish. In addition, gauge invariance implies that $\mathfrak{M}$ must vanish when either of the photon four-momenta, k_1 or k_2, vanishes. Taking four-momentum conservation into account, this gives us $n+2-1=n+1$ independent conditions on the low-energy behavior of $\mathfrak{M}$. Since for massless, neutral pions the pion-pole diagrams (tree diagrams) sum to a constant, independent of pion four-momenta, the $n+1$ conditions can be satisfied only if $\mathfrak{M}(\gamma+\gamma\to n\pi^0)$ vanishes[17] up to terms which are at least of order $(\text{momentum})^{n+1}$.

Our result for $\mathfrak{M}^{0+-}$ in Eq. (5b) disagrees with the formulas quoted by Aviv *et al.* and by Yao, both of which overlook the class of pole diagrams proportional to $F^{3\pi}$. The formula of Aviv *et al.* also has the 1 in the large round parentheses multiplying F^π replaced by $\frac{1}{3}$. In order to better understand this latter discrepancy, it is helpful to have a diagrammatic interpretation of the various terms in Eq. (5b). This is given in Fig. 2, which illustrates the lowest-order perturbation-theory contributions to $\mathfrak{M}^{000}$ and $\mathfrak{M}^{0+-}$ in the Gell-Mann–Lévy

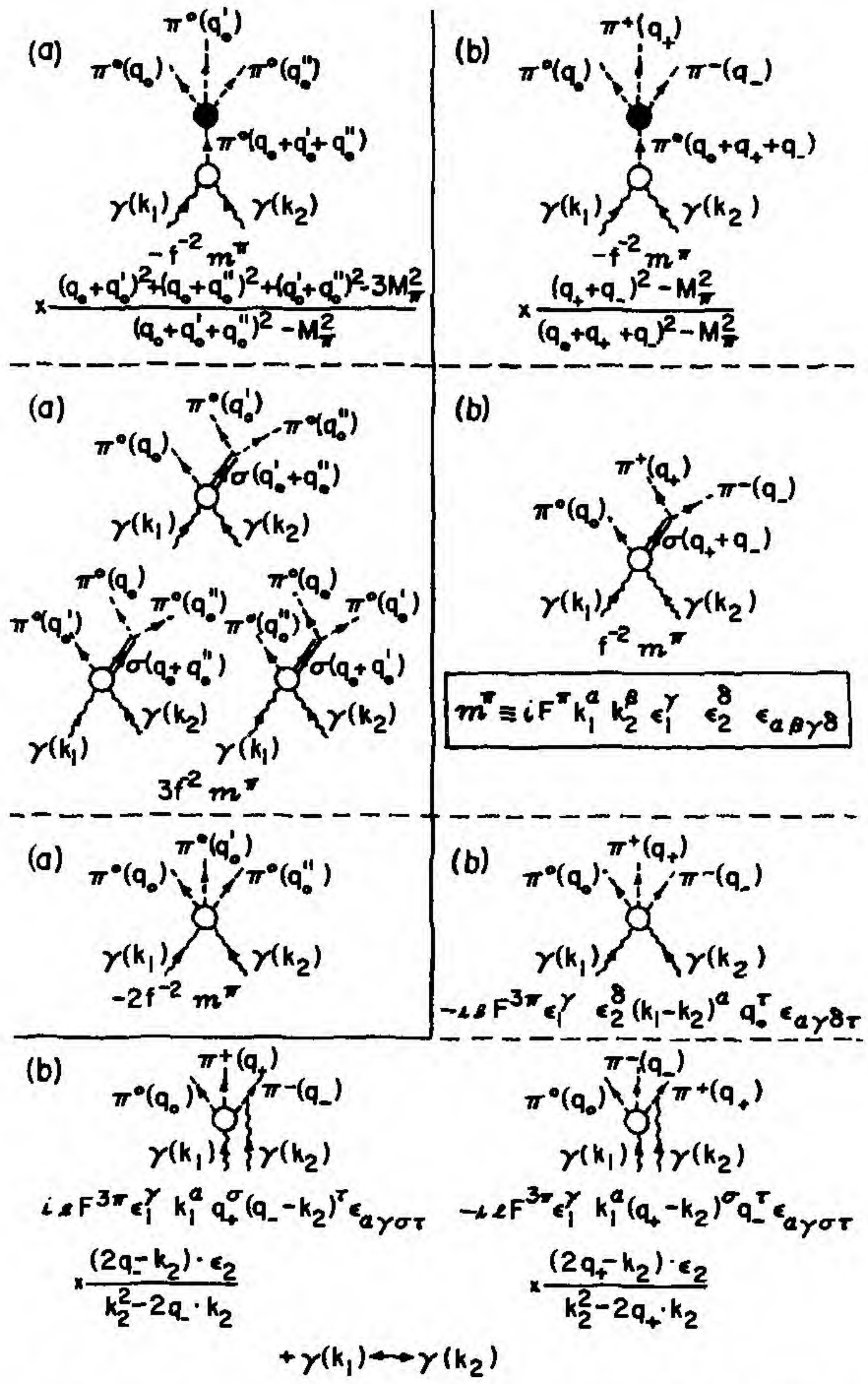


FIG. 2. Lowest-order diagrams contributing to (a) $\mathfrak{M}^{000}$ and (b) $\mathfrak{M}^{0+-}$ in the Gell-Mann – Lévy σ model. The single solid line propagating around each loop denotes the nucleon. In this order of perturbation theory, $f^{-1}=g_r/M_N$, with g_r the pion-nucleon coupling constant and with M_N the nucleon mass. {The large black dot at the four-pion vertices denotes the pion-pion scattering amplitude of Eq. (1). To lowest order in perturbation theory, this arises as the sum of a direct four-pion interaction [coming from the term $(\vec{\pi}\cdot\vec{\pi})^2$ in the σ-model Lagrangian] and of pole terms involving isoscalar σ mesons exchanged between pairs of pions.}

σ model.[18] The first and fourth rows give just the lowest-order contributions to the pole diagrams of Fig. 1. The σ-pole diagrams in the second row can clearly be represented as matrix elements of the effective Lagrangian

$$i\mathcal{L}_{\rm eff}^{\sigma}=\tfrac{1}{16}\,i f^{-2}F^{\pi}F^{\alpha\beta}F^{\gamma\delta}\epsilon_{\alpha\beta\gamma\delta}\pi^0\vec{\pi}\cdot\vec{\pi}\,, \tag{6}$$

with $F^{\alpha\beta}$ the electromagnetic field-strength tensor. As a check, we note that $\pi^0\vec{\pi}\cdot\vec{\pi}=(\pi^0)^3+2\pi^0\pi^+\pi^-$, and since the matrix element of $(\pi^0)^3$ has a Bose symmetry factor of 6, the contributions of Eq. (6) to $\mathfrak{M}^{000}$ and to $\mathfrak{M}^{0+-}$ are in the correct ratio of 3:1.

Let us turn next to the five-point functions in the third row. Aviv *et al.* assume that these are represented by the same effective-Lagrangian structure as in Eq. (6). If this were so, a five-point contribution of $-2f^{-2}\,\mathfrak{M}^{\pi}$ to $\mathfrak{M}^{000}$ would imply a corresponding contribution of $-\frac{2}{3}f^{-2}\,\mathfrak{M}^{\pi}$ to $\mathfrak{M}^{0+-}$, which would then combine with the σ-pole diagram to give a total nonpole contribution of $\frac{1}{3}f^{-2}\,\mathfrak{M}^{\pi}$. This is the origin of the $\frac{1}{3}$ in the formula of Aviv *et al.* In actual fact, however, we find that the five-point diagrams are not described by Eq. (6), but rather by the effective Lagrangian

$$i\mathcal{L}_{\text{eff}}^{5\text{-pt}} = -\tfrac{1}{2} i\, e F^{3\pi}(\partial^{\alpha}A^{\gamma})A^{\delta}\epsilon_{\alpha\gamma\delta\tau}(\partial^{\tau}\pi^{0})\vec{\pi}\cdot\vec{\pi}\,. \qquad (7)$$

Equation (7) still couples the three final pions through a pure $I=1$ state, as required by G parity. In the charged-pion case, Eq. (7) obviously leads to the five-point contribution listed in the third row of Fig. 2(b). Although not gauge-invariant by itself, this contribution combines with the pole terms in the fourth row of Fig. 2(b) (which are also not by themselves gauge-invariant) to give a gauge-invariant sum. In the neutral case, using the fact that the matrix element of $\partial^{\delta}\pi^{0}(\pi^{0})^{2}$ is $2i(q_0+q_0'+q_0'')=2i(k_1+k_2)$ and using Eq. (4) to eliminate $F^{3\pi}$ in terms of F^{π}, we find that Eq. (7) just gives the gauge-invariant contribution $-2f^{-2}\,\mathfrak{M}^{\pi}$, as required.[19] Finally, we note that while Yao obtains the correct value of 1 for the constant term in the large round parentheses multiplying F^{π}, he gets this by using an incorrect effective Lagrangian, which does not respect the $\Delta I=1$ rule, to generalize from the neutral to the charged case. The moral is that effective Lagrangians must be handled with caution. When ambiguities arise as to the form of the effective Lagrangian, they must be resolved by reference back to the basic current-algebra relations, which the effective Lagrangian is supposed to represent.[20]

ACKNOWLEDGMENTS

We wish to thank R. Aviv, R. F. Dashen, D. Gross, and R. F. Sawyer for stimulating conversations, and S. Brown and E. S. Abers for a helpful critical reading of the manuscript.

---

*Permanent address: Institute for Advanced Study, Princeton, N. J. 08540.

†Permanent address: Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, Long Island, N. Y. 11790.

‡Permanent address: Physics Department, Princeton University, Princeton, N. J. 08540.

[1]S. Brodsky, T. Kinoshita, and H. Terazawa, Phys. Rev. D 4, 1532 (1971).

[2]For a review, see S. B. Treiman, National Accelerator Laboratory Report No. THY-13 (unpublished). See also R. Aviv and R. F. Sawyer, Phys. Rev. D 4, 451 (1971).

[3]R. Aviv, N. D. Hari Dass, and R. F. Sawyer, Phys. Rev. Letters 26, 591 (1971).

[4]Tsu Yao, Phys. Letters 35B, 225 (1971).

[5]S. L. Adler and R. F. Dashen, *Current Algebras and Applications to Particle Physics* (Benjamin, New York, 1968), Chaps. 2 and 3.

[6]The reasoning is identical to that used to obtain the soft-π^0 theorem for η decay; see D. G. Sutherland, Phys. Letters 23, 384 (1966), and S. L. Adler, Phys. Rev. Letters 18, 519 (1967) for details.

[7]S. L. Adler, Phys. Rev. 177, 2426 (1969); J. S. Bell and R. Jackiw, Nuovo Cimento 60, 47 (1969); W. A. Bardeen, Phys. Rev. 184, 1848 (1969).

[8]For example, the $\gamma+\gamma\to\pi^0$ matrix element given in Eq. (2) below may be rewritten, by using four-momentum conservation, in the form $\mathfrak{M}(\gamma+\gamma\to\pi^0)=iq_0^{\alpha}k_2^{\beta}\epsilon_1^{\gamma}\epsilon_2^{\delta}\epsilon_{\alpha\beta\gamma\delta}F^{\pi}$, and so vanishes when $q_0=0$. The effect of the PCAC anomaly on this reaction is to make soft-pion calculations give $F^{\pi}\neq 0$.

[9]In fact, the diagrammatic analysis given below in Fig. 2 shows that the soft-π^0 limit of $\mathfrak{M}(\gamma+\gamma\to\pi+\pi+\pi)$ involves only axial-vector Ward identities for ring diagrams which have pseudoscalar (and in some cases scalar) vertices in addition to vector vertices and the axial-vector vertex. These Ward identities are known *not* to have anomalies; see W. A. Bardeen, Ref. 7, and R. W. Brown, C.-C. Shih, and B. L. Young, Phys. Rev. 186, 1491 (1969).

[10]Since the matrix elements in question are even functions of the external four-momenta, the error will actually be of order (momentum)^4.

[11]S. Weinberg, Phys. Rev. Letters 17, 616 (1966).

[12]We use the notation and metric conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), pp. 377–390.

[13]For a discussion of experimental evidence on the sign of F^{π}, see F. J. Gilman, Phys. Rev. 184, 1964 (1969), and S. L. Adler, in *Proceedings of the Third International Conference on High Energy Physics and Nuclear Structure, New York, 1969,* edited by S. Devons (Plenum, New York, 1970), p. 647.

[14]In a renormalizable fermion-triplet model which satisfies PCAC, the anomaly predictions for F^{π} and $F^{3\pi}$ individually are $F^{\pi}\approx-(\alpha/\pi)f^{-1}2\overline{Q}$ and $F^{3\pi}\approx-(e/4\pi^2)\times f^{-3}2\overline{Q}$. The quantity $\overline{Q}$, which is the average charge of the nonstrange triplet particles, drops out in the ratio. See S. L. Adler, Ref. 7; S. L. Adler and W. A. Bardeen, Phys. Rev. 182, 1517 (1969); and R. Aviv and A. Zee (unpublished).

[15]In the large square-bracketed terms in Eq. (5b), we have specialized to the case in which the charged pions are on the mass shell: $q_+^2=q_-^2=M_{\pi}^2$.

[16]However, one cannot conclude that $\gamma+\gamma\to\pi^0+\pi^0+\pi^0$ is suppressed relative to $\gamma+\gamma\to\pi^0+\pi^++\pi^-$. In fact, for all values of the parameter x the *threshold* value of $\mathfrak{M}^{000}$ is *three times larger* than that of $\mathfrak{M}^{0+-}$, as required by the $\Delta I=1$ rule.

[17]This generalizes the result of E. S. Abers and S. Fels, Phys. Rev. Letters 26, 1512 (1971).

[18]M. Gell-Mann and M. Lévy, Nuovo Cimento 16, 705 (1960). The σ meson in this model is pure isoscalar, and so the parameter x vanishes. A similar calculation in the σ model has been done independently by T. F. Wong (unpublished).

[19]We emphasize that this consistency check means that Eq. (4) is a model-independent result, since it is required by Eq. (5), together with the fact that the *only* two-derivative 2γ-3π couplings consistent with the $\Delta I = 1$ rule are given by Eqs. (6) and (7). For a closely related discussion, see J. Wess and B. Zumino, Phys. Letters (to be published).

[20]After this work was completed, we learned that similar results have been obtained independently by M. V. Terentiev. See M. V. Terentiev, Zh. Eksperim. i Teor. Fiz. Pis'ma v Redaktsiyu 14, 140 (1971).

VOLUME 22, NUMBER 18 PHYSICAL REVIEW LETTERS 5 MAY 1969

BREAKDOWN OF ASYMPTOTIC SUM RULES IN PERTURBATION THEORY

Stephen L. Adler and Wu-Ki Tung
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 5 March 1969)

It is shown that all of the principal results of the Bjorken-limit technique break down in perturbation theory in the "gluon" model of strong interactions.

Three years ago Bjorken[1] pointed out that the asymptotic behavior of a time-ordered product of two currents is related to equal-time commutators of the currents and their time derivatives,

$$\lim_{\substack{q_0\to i\infty \\ \vec{q}\text{ fixed}}} \int d^4x\, e^{-iq\cdot x} T(J_\mu{}^a(x) J_\nu{}^b(0)) = (iq_0)^{-1}\int d^4x\, e^{-iq\cdot x}\delta(x^0)[J_\mu{}^a(x), J_\nu{}^b(0)] + (iq_0)^{-2}\int d^4x\, e^{-iq\cdot x}\delta(x^0)[\dot{J}_\mu{}^a(x), J_\nu{}^b(0)] + O(q_0{}^{-3}), \quad (1)$$

$$\dot{J}_\mu{}^a(x) = (\partial/\partial x^0) J_\mu{}^a(x).$$

This connection has been extensively applied to the study of radiative corrections to hadronic β decay[2] and to the derivation of asymptotic sum rules[3] and asymptotic cross-section relations[4] for high energy inelastic electron and neutrino scattering. In all of these applications, it is assumed that the equal-time commutators appearing on the right-hand side of Eq. (1) are the same as the "naive commutators" obtained by straightforward use of canonical commutation relations and equations of motion. That this is a questionable assumption was pointed out by Johnson and Low,[5] who independently discovered Eq. (1). They studied this equation in a simple perturbation-theory model, in which the currents couple through a fermion triangle loop to a scalar (vector) meson. They found that in most cases the results obtained by explicit evaluation of the left-hand side of Eq. (1) differ from those calculated from naive commutators by well-defined extra terms. Because of special features of the triangle graph model, however, these extra terms do not directly invalidate the applications of Eq. (1) mentioned above.

We report here the results of a more realistic perturbation theory calculation, which shows that for commutators of space components with space components, the Bjorken limit and the naive commutator differ by terms which modify all of the principal applications of Eq. (1). We consider a simple, renormalizable model of strong interactions, consisting of an SU(3) triplet of spin-$\frac{1}{2}$ particles ψ bound by the exchange of an SU(3)-singlet massive vector "gluon." The vector current in this model is $J_\mu{}^a = \bar{\psi}\gamma_\mu\lambda^a\psi$, and the naive equal-time commutator of two vector currents is

$$\delta(x^0-y^0)[J_\mu{}^a(x), J_\nu{}^b(y)] = \delta^4(x-y)\bar{\psi}(x)C\psi(x), \quad (2)$$

$$C = \tfrac{1}{2}\{\lambda^a,\lambda^b\}(\gamma_\mu\gamma_0\gamma_\nu - \gamma_\nu\gamma_0\gamma_\mu) + \tfrac{1}{2}[\lambda^a,\lambda^b](\gamma_\mu\gamma_0\gamma_\nu + \gamma_\nu\gamma_0\gamma_\mu).$$

We wish to compare the Bjorken-limit commutator with the naive commutator, to second order in the gluon-fermion coupling constant g, in the special case in which Eqs. (1) and (2) are sandwiched between fermion states. To do this, we calculate the renormalized current-fermion scattering amplitude $\tilde{T}_{\mu\nu}{}^{ab}(p,p',q)$ and compare it, in the limit as $q_0\to i\infty$, with the renormalized vertex $\tilde{\Gamma}(C;p,p')$ of the naive commutator.[6] The scattering amplitude can be expressed in terms of the renormalized vector vertex $\tilde{\Gamma}(\gamma_\mu;p,p')$ and the renormalized fermion propagator $\tilde{S}(p)$ by

$$\tilde{T}_{\mu\nu}{}^{ab}(p,p',q) = \tilde{\Gamma}(\gamma_\mu;p,p+q)\tilde{S}(p+q)\tilde{\Gamma}(\gamma_\nu;p+q,p')\lambda^a\lambda^b + \tilde{\Gamma}(\gamma_\nu;p,p-q')\tilde{S}(p-q')\tilde{\Gamma}(\gamma_\mu;p-q',p')\lambda^b\lambda^a + B_{\mu\nu}{}^{ab}(p,p',q), \quad (3)$$

with $B_{\mu\nu}{}^{ab}(p,p',q)$ the sum of the two box diagrams illustrated in Fig. 1. We find, by explicit calcu-

lation,

$$\lim_{\substack{q_0 \to i\infty \\ \vec{q},\, p,\, p' \text{ fixed}}} \bar{T}_{\mu\nu}^{ab}(p,p',q) = q_0^{-1}[\bar{\Gamma}(C;p,p') + \Delta] + O(q_0^{-2}\ln q_0), \tag{4}$$

$$\Delta = (g^2/16\pi^2)\{2(g_{\mu\nu} - g_{\mu 0}g_{\nu 0})\gamma_0[\lambda^a,\lambda^b] + \tfrac{3}{2}(\gamma_\nu\gamma_0\gamma_\mu - \gamma_\mu\gamma_0\gamma_\nu)\{\lambda^a,\lambda^b\}\}.$$

We see that the Bjorken-limit commutator and the naive commutator differ by the term labeled Δ, which is well defined and finite. We note that Δ vanishes when $\mu = 0$ or $\nu = 0$, indicating that for the time-time and time-space commutators, the Bjorken limit and the naive commutator agree. This result can be independently deduced from the usual on-shell Ward identity

$$q^\mu \bar{T}_{\mu\nu}^{ab}(p,p',q) = \bar{\Gamma}([\lambda^a,\lambda^b]\gamma_\nu; p,p'); \tag{5}$$

the consistency between Eq. (4) and Eq. (5) provides a convenient check on the calculation leading to Eq. (4). When one or both currents J_μ^a, J_ν^b is replaced by the corresponding axial-vector current $J_\mu^{5a} = \bar{\psi}\gamma_\mu\gamma_5\lambda^a\psi$, J_ν^{5b}, a formula like Eq. (4) holds, with the appropriate change in C and with Δ modified as follows:

$$J_\mu^a \to J_\mu^{5a} \Longleftrightarrow \Delta \to -\gamma_5\Delta, \quad J_\nu^b \to J_\nu^{5b} \Longleftrightarrow \Delta \to \Delta\gamma_5, \quad J_\mu^a \to J_\mu^{5a}, J_\nu^b \to J_\nu^{5b} \Longleftrightarrow \Delta \to -\gamma_5\Delta\gamma_5 = \Delta. \tag{6}$$

One may wonder whether our definition of $\bar{\Gamma}(C;p,p')$ could be changed by a finite rescaling in such a way as to absorb the term Δ. However, since $\gamma_\mu\gamma_0\gamma_\nu + \gamma_\nu\gamma_0\gamma_\mu \propto g_{\mu 0}\gamma_\nu + g_{\nu 0}\gamma_\mu - g_{\mu\nu}\gamma_0$ and since $\gamma_\mu\gamma_0\gamma_\nu - \gamma_\nu\gamma_0\gamma_\mu \propto \epsilon_{\mu 0\nu\lambda}\gamma^\lambda\gamma_5$, the vertex $\bar{\Gamma}(C;p,p')$ is a linear combination of vector and axial-vector vertices. Therefore, the normalization of this vertex is completely fixed by the time-component current algebra and Lorentz covariance, and rescaling is not permitted.

In addition to studying the q_0^{-1} term in Eq. (1), we have also calculated the q_0^{-2} term in the special case considered by Callan and Gross.[4] Specializing to forward scattering ($p = p'$, $a = b$) and spin averaging, we find

$$\lim_{p_0 \to \infty} \lim_{q_0 \to i\infty} m p_0^{-2} q_0^{2} \tfrac{1}{4}\operatorname{tr}\left[\left(\frac{\gamma\cdot p + m}{2m}\right)\bar{T}_{ij}^{aa}(p,p,q)\right]$$
$$= -2(\delta^{ij} - \hat{p}^i\hat{p}^j)(\lambda^a)^2 + \frac{g^2}{6\pi^2}[2(\ln q_0^{\,2} + \text{const})(\delta^{ij} - \hat{p}^i\hat{p}^j) + \hat{p}^i\hat{p}^j](\lambda^a)^2. \tag{7}$$

The presence of $\ln q_0^{\,2}$ on the right-hand side of Eq. (7) indicates that the expression of Eq. (1) cannot, strictly speaking, be carried out to order q_0^{-2}, and that the coefficient $\langle p|\delta(x^0)[\dot{J}_i^a(x), J_j^a(0)]|p\rangle$ of the q_0^{-2} term is logarithmically divergent.[7] Using naive commutators to evaluate this coefficient, Callan and Gross concluded that the double limit on the left-hand side of Eq. (7) should be proportional to the transverse tensor $\delta^{ij} - \hat{p}^i\hat{p}^j$. The presence of the additional term $(g^2/6\pi^2)\hat{p}^i\hat{p}^j(\lambda^a)^2$ in Eq. (7) indicates that their conclusion fails in perturbation theory.

We next indicate how the various applications of the Bjorken-limit technique are modified by our results.

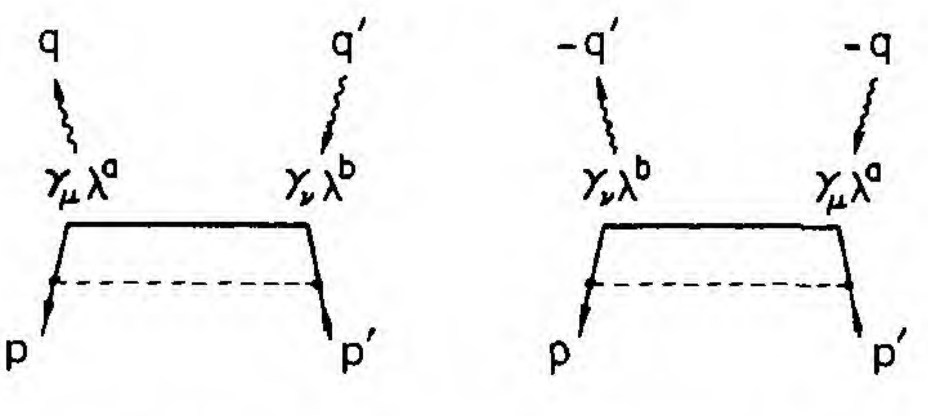


FIG. 1. Box diagrams contributing to $B_{\mu\nu}^{ab}$. The dashed line denotes the virtual gluon.

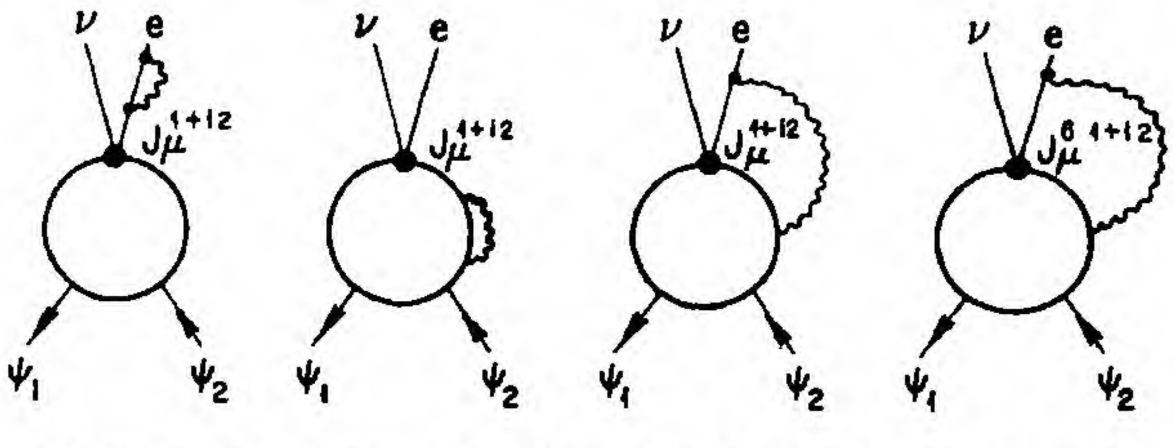


FIG. 2. Diagrams for the radiative corrections to the vector β transition. The wavy line denotes the virtual photon.

VOLUME 22, NUMBER 18 PHYSICAL REVIEW LETTERS 5 MAY 1969

(i) Radiative corrections to β decay.[2]—We consider the vector β transition between the fermions ψ_1 and ψ_2. We introduce a cutoff Λ^2 and calculate the divergent part of the radiative corrections to this process, described by the diagrams of Fig. 2. Using the time-component current algebra alone, it has been shown that the first three diagrams in Fig. 2 sum to a universal, structure-independent fractional change in the decay amplitude $\delta M/M = (3\alpha/8\pi)\ln\Lambda^2$. The divergent part of the fourth diagram in Fig. 2 can be evaluated to order g^2 from Eqs. (4) and (6), giving $\delta M/M = (3\alpha/8\pi)2\bar{Q}\ln\Lambda^2(1-3g^2/16\pi^2)$, with $\bar{Q}$ the average charge of the doublet $\psi_{1,2}$. The term proportional to g^2 comes, of course, from the isospin-symmetric term in Δ. The total divergent part of the radiative correction is thus, to order g^2,

$$(\delta M/M)_{\text{total}} = (3\alpha/8\pi)\ln\Lambda^2[1+2\bar{Q}(1-3g^2/16\pi^2)]. \tag{8}$$

We see that the choice $\bar{Q} = -\frac{1}{2}$, which removes the divergence to lowest order in g^2, still leaves a residual divergence in second order.

(ii) Asymptotic sum rules and cross-section relations.[3,4]—We introduce the variable $\omega = -q^2/p\cdot q$ and define the spectral functions $W_1(\omega,q^2)$ and $W_2(\omega,q^2)$ by

$$\frac{\operatorname{disc}_\omega}{-2\pi i}\frac{1}{4}\operatorname{tr}\left[\left(\frac{\gamma\cdot p+m}{2m}\right)\tilde{T}_{\mu\nu}{}^{ab}(p,p,q)\right] = \lambda^a\lambda^b[W_1(\omega,q^2)(-g_{\mu\nu}+q_\mu q_\nu/q^2)$$
$$+W_2(\omega,q^2)(p_\mu - p\cdot qq_\mu/q^2)(p_\nu - p\cdot qq_\nu/q^2)],\quad \omega>0. \tag{9}$$

In terms of these spectral functions, the asymptotic formula of Eq. (7) may be rewritten as the sum rule

$$\lim_{q^2\to-\infty}(\cdots)(\delta^{ij}-\hat{p}^i\hat{p}^j)+\frac{g^2}{6\pi^2}\hat{p}^i\hat{p}^j = \lim_{q^2\to-\infty} 2\int_0^2\omega d\omega[mW_1(\delta^{ij}-\hat{p}^i\hat{p}^j)+(mW_1+q^2mW_2/\omega^2)\hat{p}^i\hat{p}^j], \tag{10}$$

first obtained by Callan and Gross. As these authors note, the quantity $mW_1+q^2mW_2/\omega^2$ is positive definite, differing only by positive factors from $-q^2\sigma_L(\omega,q^2)$, with σ_L the longitudinal electroproduction cross section. Thus the presence of $\hat{p}^i\hat{p}^j$ in Eq. (7) implies that, in the quark model, $q^2\sigma_L(\omega,q^2)$ does not vanish asymptotically, in disagreement with the conclusion of Callan and Gross. In a similar fashion, the SU(3)-antisymmetric part of Eq. (4) leads to the asymptotic sum rule

$$1-\frac{g^2}{8\pi^2} = \lim_{q^2\to-\infty} -2\int_0^2 d\omega\, mW_1. \tag{11}$$

Apart from the term $g^2/8\pi^2$ on the left-hand side, which comes from the SU(3)-antisymmetric part of Δ, Eq. (11) is the backward-neutrino-scattering asymptotic sum rule of Bjorken.[3] The modification in the left-hand side of Eq. (11) is closely related to the nonvanishing of $q^2\sigma_L(\omega,q^2)$. To see this, we write down the usual fixed-q^2, time-component algebra sum rule[8]

$$1 = 2\int_0^2 d\omega\, q^2 mW_2/\omega^2 \tag{12}$$

and subtract it from Eq. (11), giving[9]

$$-g^2/8\pi^2 = \lim_{q^2\to-\infty} -2\int_0^2 d\omega(mW_1+q^2mW_2/\omega^2). \tag{13}$$

Thus the SU(3)-antisymmetric term in Δ and the $\hat{p}^i\hat{p}^j$ term in Eq. (7) are basically the same phenomenon. As an additional check on our arithmetic, we have calculated W_1 and W_2 directly, giving $mW_1 + q^2mW_2/\omega^2 = g^2\omega/32\pi^2$, in agreement with Eqs. (10) and (13).

We have also studied the scalar (pseudoscalar) gluon model in perturbation theory, and find effects similar to those reported here. Full details of the calculations, and further discussion, will be published elsewhere.

We wish to thank W. A. Bardeen and S. B. Treiman for helpful discussions, and Dr. Carl Kaysen for the hospitality of the Institute for Advanced Study. After this work was completed, we learned that R. Jackiw and G. Preparata had also discovered the breakdown of the Callan-Gross result in perturbation theory.

VOLUME 22, NUMBER 18 PHYSICAL REVIEW LETTERS 5 MAY 1969

Note added in proof.—(i) We have been informed by J. D. Bjorken of a related paper by A. I. Vainshtein and B. L. Ioffe {Zh. Eksperim. i Teor. Fiz.—Pis'ma Redakt. 6, 917 (1967) [translation: Soviet Phys.-JETP Letters 6, 341 (1967)]}. We will discuss this work in our detailed paper. (ii) In the case when one current is an axial-vector current [the first two lines of Eq. (6)], we have omitted an SU(3)-singlet contribution to the Bjorken limit coming from the triangle diagram discussed by Johnson and Low. Addition of this piece does not alter any of our conclusions.

---

[1]J. D. Bjorken, Phys. Rev. 148, 1467 (1966).

[2]For references, see G. Preparata and W. I. Weisberger, Phys. Rev. 175, 1965 (1968).

[3]For a survey, see lectures by J. D. Bjorken, in Selected Topics in Particle Physics, Proceedings of the International School of Physics "Enrico Fermi," Course XLI, edited by J. Steinberger (Academic Press, Inc., New York, 1968).

[4]C. G. Callan and D. J. Gross, Phys. Rev. Letters 22, 156 (1969).

[5]K. Johnson and F. E. Low, Progr. Theoret. Phys. (Kyoto) Suppl. Nos. 37-38, 74 (1966).

[6]The renormalized vertex $\tilde{\Gamma}(C;p,p')$ is obtained from the unrenormalized vertex $\Gamma(C;p,p')$ by multiplying by the fermion wave-function renormalization constant Z_2, with no further finite rescalings.

[7]The term $\ln q_0^2$ is present when p_0 is finite and is not a result of the additional $p_0 \to \infty$ limit.

[8]See S. L. Adler and R. F. Dashen, Current Algebras (W. A. Benjamin, Inc., New York, 1968), Chap. 4.

[9]This connection was first noted by F. J. Gilman, Phys. Rev. 167, 1365 (1968).

PHYSICAL REVIEW D VOLUME 1, NUMBER 10 15 MAY 1970

Bjorken Limit in Perturbation Theory

STEPHEN L. ADLER AND WU-KI TUNG
Institute for Advanced Study, Princeton, New Jersey 08540

(Received 29 October 1969)

We present detailed calculations illustrating the breakdown of the Bjorken limit in perturbation theory, in the "gluon" model of strong interactions. To second order in the gluon-ferrmion coupling constant in the scalar, pseudoscalar, and vector coupling models, we calculate the Bjorken-limit commutator of a pair of currents of arbitrary (vector, axial-vector, scalar, pseudoscalar, tensor) type. To fourth order in the coupling, in the scalar- and pseudoscalar-gluon models, we determine the leading logarithmic behavior of the SU_3-antisymmetric part of the vector-vector commutator. In the body of the paper we present the main results and discuss their various features and implications. The computational details are relegated to two appendices.

I. HISTORICAL INTRODUCTION

EQUAL-TIME current commutators have come to play a central role in particle physics. In his famous papers of 1961 and 1964, Gell-Mann[1] proposed that the time components of the vector and axial-vector octet currents satisfy a simple $SU_3 \otimes SU_3$ algebra. The exploitation of this postulate by the "infinite-momentum" and "low-energy theorem" methods has led to important predictions, which agree well with experiment.[2] The beauty of these "classical" current-algebra methods is that they depend only on the postulated commutation relations together with such weak dynamical assumptions as pion-pole dominance and unsubtracted dispersion relations. They are independent of more detailed (and therefore, more dubious) dynamical assumptions. The experimental successes thus provide a strong argument that any future theory of the hadrons must incorporate the $SU_3 \otimes SU_3$ time-component current algebra.

This requirement, of course, does not uniquely specify a model of the hadrons—there are many possible field-theoretic models which satisfy the Gell-Mann hypothesis. In an attempt to narrow the selection, attention has been turned recently to the study of the space-component–space-component commutators, which can be used to distinguish between models which have the same time-component algebra. The problem of finding experimental tests of the space-space algebra is made difficult by the fact that the "classical" current-algebra methods of infinite-momentum limits and low-energy theorems cannot be made to apply in this case. However, in 1966 Bjorken[3] pointed out that the asymptotic behavior of a time-ordered product of two currents is simply related to the equal-time commutator of the currents,

$$\lim_{q_0\to i\infty;\,q\ \text{fixed}} \int d^4x\, e^{iq\cdot x}T(J_{(1)}(x)J_{(2)}(0))$$
$$= iq_0^{-1}\int d^4x\, e^{iq\cdot x}\delta(x^0)[J_{(1)}(x),J_{(2)}(0)]+O(q_0^{-2}). \quad (1)$$

Equation (1) has been extensively applied to the study of space-space current commutators, leading to a new class of *asymptotic sum rules*.[4] These sum rules have testable experimental consequences in inelastic electron and neutrino scattering reactions and important implications in the theory of radiative corrections to hadronic β decay.

In all of the applications of Eq. (1), an important assumption is made: It is *assumed* that the equal-time commutator appearing on the right-hand side of Eq. (1) is the same as the "naive commutator" obtained by straightforward use of canonical commutation relations and equations of motion. That this is a questionable assumption was pointed out by Johnson and Low,[5] who independently discovered Eq. (1). They studied this equation in a simple perturbation-theory model, in which the currents couple through a fermion triangle loop to a scalar, pseudoscalar, or vector meson. They found that in most cases the results obtained by explicit evaluation of the left-hand side of Eq. (1) differ from those calculated from naive commutators by well-defined extra terms. Because of special features of the triangle graph model, however, these extra terms did not directly invalidate the applications of Eq. (1) mentioned above.

Recently, we have reported a more realistic perturba-

[1] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962); Physics **1**, 63 (1964).
[2] For a survey, see S. L. Adler and R. F. Dashen, *Current Algebras* (Benjamin, New York, 1968).
[3] J. D. Bjorken, Phys. Rev. **148**, 1467 (1966).
[4] For a survey, see lectures by J. D. Bjorken, in *Selected Topics in Particle Physics, Proceedings of the International School of Physics "Enrico Fermi," Course XLI*, edited by J. Steinberger (Academic, New York, 1968).
[5] K. Johnson and F. E. Low, Progr. Theoret. Phys. (Kyoto) Suppl. Nos. **37-38**, 74 (1966). Important early work on the validity of the Bjorken limit, in the context of the Lee model, has also been done by J. S. Bell, Nuovo Cimento **47A**, 616 (1967).

tion-theory calculation,[6] which showed that for commutators of space components with space components, the Bjorken limit and the naive commutator do differ by terms which modify all of the principal applications of Eq. (1). In other words, *asymptotic sum rules derived from the naive space-space commutators fail in perturbation theory.* One is, of course, still free to postulate that nonperturbative effects conspire to make the asymptotic sum rules valid when all orders of perturbation theory are summed, but the need for this assumption means that asymptotic sum rules do not just give a test of the space-space algebra, but involve deep dynamical considerations as well.

In our previous work, we considered only vector and axial-vector current commutators in the quark model with a massive vector "gluon," to second order in the gluon-fermion coupling constant g_r.[7] In the present paper, we extend our results to arbitrary (vector, axial-vector, scalar, pseudoscalar, tensor) currents in the quark models with vector-, scalar-, or pseudoscalar-coupled gluon. Working to second order in g_r, we obtain results analogous to those found previously in the more restricted case. In addition, for the vector-vector commutator in the scalar- and pseudoscalar-gluon models, we obtain the leading logarithmic part of the g_r^4 term. In Sec. II we summarize our results and in Sec. III we discuss briefly their significance. To facilitate reading, all computational details are relegated to Appendices.

II. RESULTS

We consider a simple, renormalizable model of the strong interactions, consisting of an SU_3 triplet of spin-$\frac{1}{2}$ particles ψ bound by the exchange of an SU_3-singlet massive "gluon." We assume that the gluon couples to the fermions by either scalar, pseudoscalar, or vector coupling. In order to treat simultaneously commutators involving vector (axial-vector, scalar, ...) currents, we introduce the abbreviated notation

$$J_{(1)}=\bar{\psi}\gamma_{(1)}\psi, \quad J_{(2)}=\bar{\psi}\gamma_{(2)}\psi,$$
$$\gamma_{(1)}=\gamma_\mu\lambda^a(\gamma_\mu\gamma_5\lambda^a, \lambda^a, \ldots), \tag{2}$$
$$\gamma_{(2)}=\gamma_\nu\lambda^b(\gamma_\nu\gamma_5\lambda^b, \lambda^b, \ldots),$$

according to whether the first or second current is a vector (axial-vector, scalar, ...) current. The naive equal-time commutator of the two currents is

$$\delta(x^0-y^0)[J_{(1)}(x), J_{(2)}(y)]=\delta^4(x-y)\bar{\psi}(x)C\psi(x),$$
$$C=\gamma_0[\gamma_0\gamma_{(1)}, \gamma_0\gamma_{(2)}]=\gamma_{(1)}\gamma_0\gamma_{(2)}-\gamma_{(2)}\gamma_0\gamma_{(1)}. \tag{3}$$

We wish to compare the Bjorken-limit commutator with the naive commutator, in the special case in which Eqs. (1) and (3) are sandwiched between fermion states. To do this, we calculate the renormalized current-fermion scattering amplitude $\tilde{T}_{(1)(2)}^*(p, p', q)$ in the limit $q_0\to i\infty$, and compare the coefficient of the q_0^{-1} term with the renormalized vertex $\tilde{\Gamma}(C; p, p')$ of the naive commutator. The asterisk on $\tilde{T}_{(1)(2)}^*$ indicates that it is the full covariant scattering amplitude, which differs from the renormalized T product, $\tilde{T}_{(1)(2)}(p, p', q)$, by a "seagull" term $\tilde{\sigma}_{(1)(2)}(p, p', q)$ which is a polynomial in q_0,

$$\tilde{T}_{(1)(2)}^*(p, p', q)=\tilde{\sigma}_{(1)(2)}(p, p', q)+\tilde{T}_{(1)(2)}(p, p', q). \tag{4}$$

Identity of the Bjorken limit and naive commutators would mean that

$$\lim_{q_0\to i\infty;\, q, p, p' \text{ fixed}} \tilde{T}_{(1)(2)}(p, p', q)=q_0^{-1}\tilde{\Gamma}(C; p, p') +O(q_0^{-2}\ln q_0). \tag{5}$$

In the calculation which follows, we test the validity of Eq. (5) in perturbation theory.[8]

A. Second Order

To second order in the gluon-fermion coupling constant g_r, there are two classes of diagrams which contribute to $\tilde{T}_{(1)(2)}^*$. The diagrams of the first class, illustrated in Fig. 1, consist of the lowest-order current-

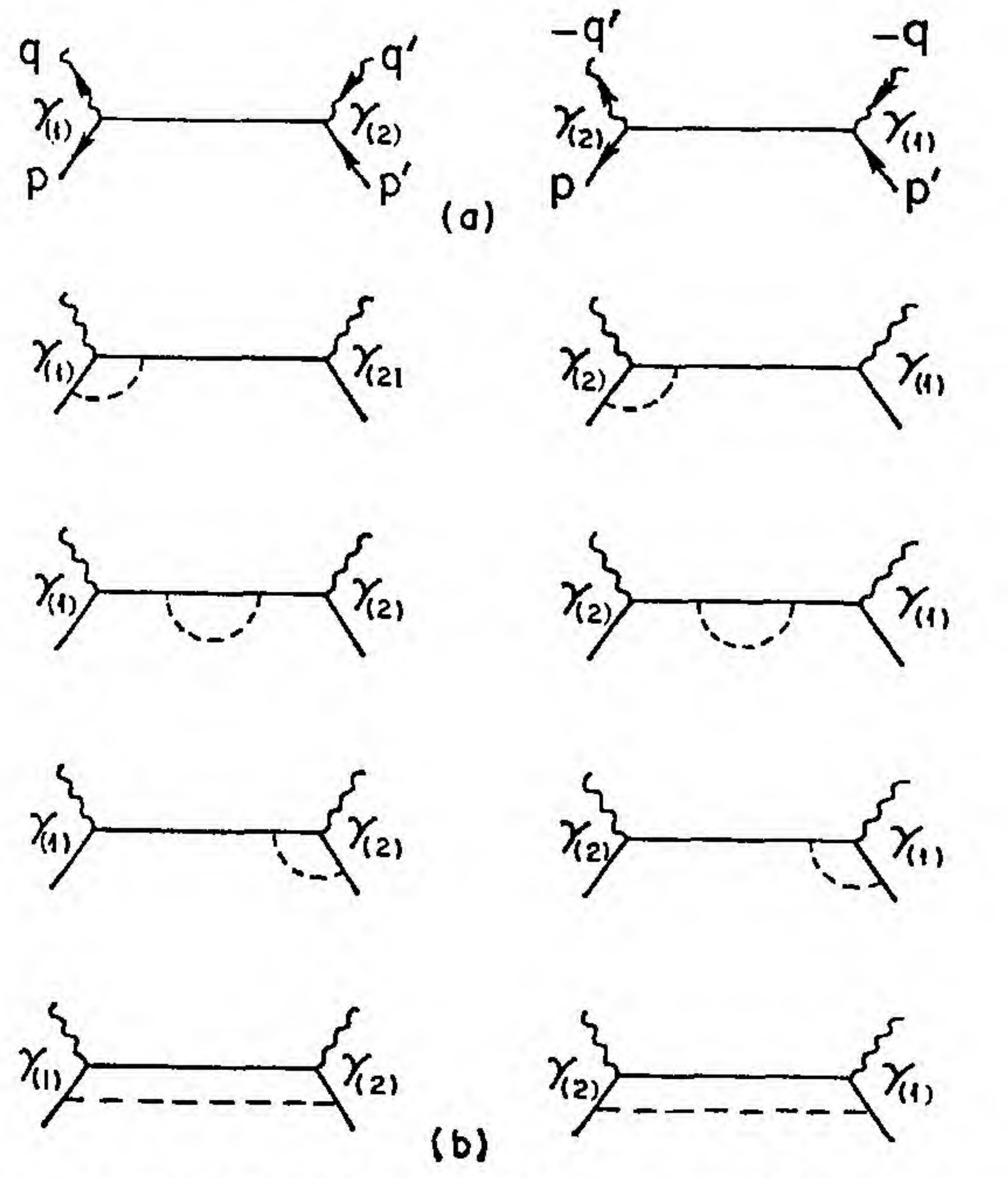


FIG. 1. (a) Lowest-order current-fermion scattering diagrams. (b) Diagrams obtained from the lowest-order ones by insertion of a single virtual gluon.

[6] S. L. Adler and W.-K. Tung, Phys. Rev. Letters 22, 978 (1969). See also R. Jackiw and G. Preparata, *ibid.* 22, 975 (1969), who have independently arrived at similar conclusions.

[7] In Ref. 6 we denoted the coupling constant g_r by g. In the present work, g will always indicate a gluon (or its four-momentum).

[8] A general discussion of the mechanism responsible for Bjorken-limit breakdown has been given by W.-K. Tung, Phys. Rev. 188, 2404 (1969). See also R. Jackiw and G. Preparata, *ibid.* 185, 1929 (1969).

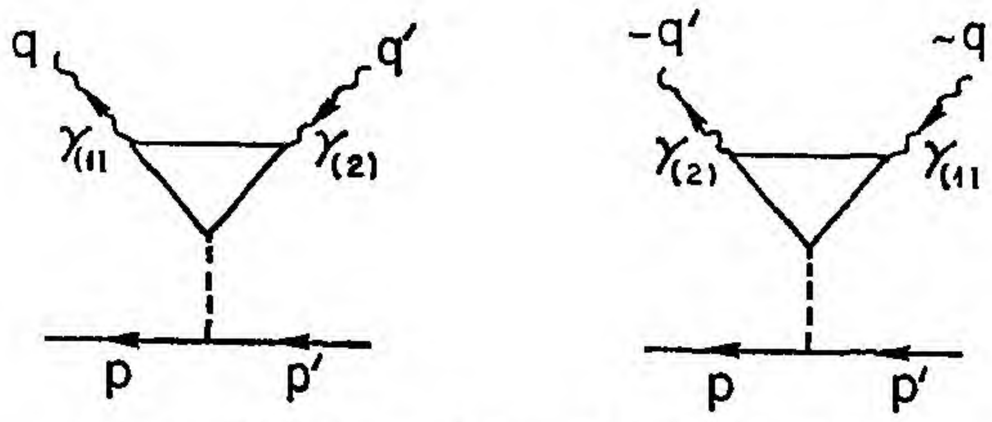


FIG. 2. Diagrams containing fermion triangles.

fermion diagrams and the second-order diagrams obtained from the lowest-order ones by insertion of a single virtual gluon. The diagrams of the second class, illustrated in Fig. 2, involve a fermion triangle diagram. We denote the contributions of these two classes to $\tilde{T}_{(1)(2)}{}^*$ by $\tilde{T}_{(1)(2)}{}^{*\mathrm{Compt}}$ and $\tilde{T}_{(1)(2)}{}^{*\mathrm{triang}}$, respectively.

The first-class diagrams are evaluated by the standard technique of regulating the gluon propagator with a regulator of mass λ, which defines an *unrenormalized* amplitude $T_{(1)(2)}{}^{*\mathrm{Compt}}$. To get the *renormalized* amplitude, one multiplies by the external fermion wave-function renormalization constant Z_2 and takes the limit $\lambda\to\infty$,

$$\tilde{T}_{(1)(2)}{}^{*\mathrm{Compt}}=\lim_{\lambda\to\infty} Z_2 T_{(1)(2)}{}^{*\mathrm{Compt}}. \tag{6}$$

In certain cases, as discussed below, this limit diverges logarithmically; in these cases, we take λ to be finite but very large, dropping terms which vanish as $\lambda\to\infty$ but retaining all terms which are proportional to $\ln\lambda^2$. The renormalized vertex

$$\tilde{\Gamma}(C;p,p')=\lim_{\lambda\to\infty} Z_2\Gamma(C;p,p')$$

is calculated by the same techniques from the diagram of Fig. 3. Finally, we take the limit $q_0\to i\infty$ in our expression for $\tilde{T}_{(1)(2)}{}^{*\mathrm{Compt}}$ and compare with $\tilde{\Gamma}(C;p,p')$, giving the results

$$\tilde{\sigma}_{(1)(2)}{}^{\mathrm{Compt}}(p,p',q)=0, \tag{7a}$$

$$\begin{aligned}\lim_{q_0\to i\infty;\,\mathbf{q},\,\mathbf{p},\,\mathbf{p}'\ \mathrm{fixed}} &\tilde{T}_{(1)(2)}{}^{*\mathrm{Compt}}(p,p',q)\\ &=\lim_{q_0\to i\infty;\,\mathbf{q},\,\mathbf{p},\,\mathbf{p}'\ \mathrm{fixed}} \tilde{T}_{(1)(2)}{}^{\mathrm{Compt}}(p,p',q)\\ &=q_0^{-1}[\tilde{\Gamma}(C;p,p')+\Delta^{\mathrm{Compt}}]+O(q_0^{-2}\ln q_0),\end{aligned} \tag{7b}$$

$$\begin{aligned}\Delta^{\mathrm{Compt}}=(g_r^2/32\pi^2)\{&\ln(\lambda^2/|q_0|^2)[-\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_0\boldsymbol{\gamma}\gamma_0\gamma_{(2)}\\ &+\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(1)}\gamma_0\gamma_{(2)}\gamma^\tau\boldsymbol{\gamma}\\ &-\tfrac{1}{2}\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_\tau\gamma_{(2)}\gamma^\tau\boldsymbol{\gamma}-\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(1)}\gamma^\tau\boldsymbol{\gamma}\gamma_0\gamma_{(2)}]\\ &-\tfrac{3}{2}\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_0\boldsymbol{\gamma}\gamma_0\gamma_{(2)}-\tfrac{1}{2}\boldsymbol{\gamma}\gamma_0\gamma_{(1)}\gamma_0\gamma_{(2)}\gamma_0\boldsymbol{\gamma}\\ &+\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_0\gamma_{(2)}\gamma_0\boldsymbol{\gamma}+\boldsymbol{\gamma}\gamma_0\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_0\gamma_{(2)}\\ &-\tfrac{1}{4}\gamma_{(1)}\gamma_0\boldsymbol{\gamma}\gamma_\tau\gamma_{(2)}\gamma^\tau\boldsymbol{\gamma}-\tfrac{1}{4}\boldsymbol{\gamma}\gamma_\tau\gamma_{(1)}\gamma^\tau\boldsymbol{\gamma}\gamma_0\gamma_{(2)}\\ &+\tfrac{1}{4}\boldsymbol{\gamma}[\gamma_\tau\gamma_{(1)}\gamma^\tau\gamma_{(2)}\gamma_0+\gamma_0\gamma_{(1)}\gamma_\tau\gamma_{(2)}\gamma^\tau]\boldsymbol{\gamma}\\ &-(1)\leftrightarrow(2)\}.\end{aligned} \tag{7c}$$

In Eq. (7), the notation $\boldsymbol{\gamma}\cdots\boldsymbol{\gamma}$ is a shorthand for $1\cdots1$ in the scalar-gluon case, $i\gamma_5\cdots i\gamma_5$ in the pseudoscalar-gluon case, and $(-\gamma_\rho)\cdots\gamma^\rho$ in the vector-gluon case. Some details of the calculation leading to Eq. (7) are given in Appendix A.

The second class diagrams (Fig. 2) have been calculated by Johnson and Low.[5] In our model, which has only SU_3-singlet gluons, these diagrams contribute only to the SU_3-singlet part of the commutator. Taking the Bjorken limit, and comparing with the bubble diagram contributions to $\tilde{\Gamma}(C;p,p')$ illustrated in Fig. 4, Johnson and Low find

$$\begin{aligned}&\lim_{q_0\to i\infty;\,\mathbf{q},\,\mathbf{p},\,\mathbf{p}'\ \mathrm{fixed}} \tilde{T}_{(1)(2)}{}^{*\mathrm{triang}}(p,p',q)\\ &=\tilde{\sigma}_{(1)(2)}{}^{\mathrm{triang}}(p,p',q)+\lim_{q_0\to i\infty}\tilde{T}_{(1)(2)}{}^{\mathrm{triang}}(p,p',q)\\ &=\tilde{\sigma}_{(1)(2)}{}^{\mathrm{triang}}(p,p',q)+q_0^{-1}[\tilde{\Gamma}(C;p,p')^{\mathrm{bubble}}+\Delta^{\mathrm{triang}}]\\ &\qquad+O(q_0^{-2}\ln q_0).\end{aligned} \tag{8}$$

We will not exhibit the detailed form of Δ^{triang}, but only remark that *in all cases* Δ^{triang} vanishes when the three-momenta $\mathbf{q}$ and $\mathbf{q}'=\mathbf{q}+\mathbf{p}-\mathbf{p}'$ associated with the currents $J_{(1)}$ and $J_{(2)}$ vanish,

$$\Delta^{\mathrm{triang}}|_{\mathbf{q}=\mathbf{q}'=0}=0. \tag{9}$$

[Equation (9) is true when the triplet of fermions ψ are degenerate in mass. Johnson and Low[5] also discuss the effect of mass splittings.] Thus, for the physically interesting case of the commutator of spatially integrated currents, the entire answer is given by Eq. (7). No cancellation between the SU_3-singlet part of Δ^{Compt} and Δ^{triang} is possible, and we conclude that the Bjorken limit and the naive commutator in our models differ in second-order perturbation theory.

To make contact with our previous work and with our fourth-order results, it is useful to write out two special cases of Eq. (7). We consider the commutator of two vector currents, taking $\gamma_{(1)}=\gamma_\mu\lambda^a$, $\gamma_{(2)}=\gamma_\nu\lambda^b$. In the vector-gluon case we find

$$\begin{aligned}\Delta^{\mathrm{Compt}}=(g_r^2/16\pi^2)\{&2(g_{\mu\nu}-g_{\mu0}g_{\nu0})\gamma_0[\lambda^a,\lambda^b]\\ &+\tfrac{3}{2}(\gamma_\nu\gamma_0\gamma_\mu-\gamma_\mu\gamma_0\gamma_\nu)\{\lambda^a,\lambda^b\}\},\end{aligned} \tag{10}$$

in agreement with the result which we have reported in Ref. 6. In the scalar- and pseudoscalar-gluon cases we find

$$\begin{aligned}\Delta^{\mathrm{Compt}}=(g_r^2/16\pi^2)\{&(g_{\mu\nu}-g_{\mu0}g_{\nu0})\gamma_0[\lambda^a,\lambda^b]\\ &-\tfrac{1}{2}(\gamma_\nu\gamma_0\gamma_\mu-\gamma_\mu\gamma_0\gamma_\nu)\{\lambda^a,\lambda^b\}[\ln(\lambda^2/|q_0|^2)-1]\}.\end{aligned} \tag{11}$$

B. Fourth Order

To fourth order in g_r, the number of diagrams contributing to $\tilde{T}_{(1)(2)}{}^*$ is so large that a direct calculation

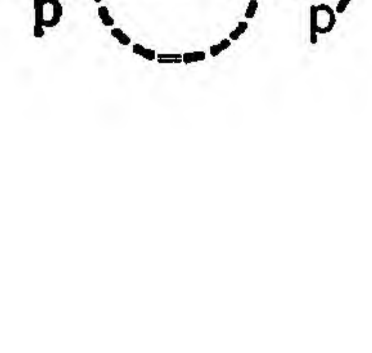


FIG. 3. Second-order correction to the vertex of the naive commutator C.

of the Bjorken limit, in analogy with our treatment of the second-order case, is prohibitively complicated. However, unitarity implies that the part of Δ proportional to $[\lambda^a, \lambda^b]$, and independent of the three-momenta $\mathbf{q}$, $\mathbf{q}'$, $\mathbf{p}$, and $\mathbf{p}'$ and of the fermion mass m, is related to an integral over the longitudinal current-fermion inelastic total cross section.[6,9] Applying this connection to the commutator of vector currents in the scalar- and pseudoscalar-gluon cases, we have calculated the leading logarithmic contribution to the $[\lambda^a, \lambda^b]$ term in fourth order, with the result

$$\begin{aligned}\Delta=(g_{\mu\nu}-g_{\mu 0}g_{\nu 0})\gamma_0[\lambda^a,\lambda^b][(g_r^2/16\pi^2)+7(g_r^2/16\pi^2)^2\\ \times\ln(|q_0|^2)+g_r^4\times\text{const}]+(\text{terms symmetric in } a,\, b)\\ +(\text{terms proportional to } \mathbf{q},\, \mathbf{q}',\, \mathbf{p},\, \mathbf{p}',\text{ and } m).\quad(12)\end{aligned}$$

Details of the unitarity relation and of the total cross-section calculation are outlined in Appendix B.

III. DISCUSSION

We proceed next to discuss a number of features of our results of Eqs. (7) and (10)-(12).

1. We begin by noting that to second order in g_r^2, Δ^{Compt} contains terms $\ln(\lambda^2/|q_0|^2)$ which diverge logarithmically both in the Bjorken limit $q_0\to i\infty$ and in the infinite-cutoff limit $\lambda\to\infty$. It is easy to see that the $\ln\lambda^2$ divergences result from a mismatch between the multiplicative factors needed to make $T_{(1)(2)}{}^{*\text{Compt}}(p,\,p',\,q)$ and $\Gamma(C;\,p,\,p')$ finite (i.e., $\ln\lambda^2$ independent). As we recall, the *renormalized* quantities $\tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,\,p',\,q)$ and $\tilde{\Gamma}(C;\,p,\,p')$ are obtained from $T_{(1)(2)}{}^{*\text{Compt}}(p,\,p',\,q)$ and $\Gamma(C;\,p,\,p')$ by multiplying by the *wave-function renormalization* Z_2 and taking the limit $\lambda\to\infty$, keeping any residual $\ln\lambda^2$ dependence. On the other hand, the *finite* quantities $T_{(1)(2)}{}^{*\text{Compt}}(p,\,p',\,q)^{\text{finite}}$ and $\Gamma(C;\,p,\,p')^{\text{finite}}$ are obtained by multiplying by appropriate vertex and propagator renormalization factors which completely remove the $\ln\lambda^2$ dependence,

$$\begin{aligned}\Gamma(C;p,p')^{\text{finite}}&=Z(C)\Gamma(C;p,p'),\\ T_{(1)(2)}{}^{*\text{Compt}}(p,p',q)^{\text{finite}}&=Z(\gamma_{(1)})Z(\gamma_{(2)})Z_2^{-1}\\ &\quad\times T_{(1)(2)}{}^{*\text{Compt}}(p,p',q).\end{aligned}\quad(13)$$

In general, the vertex renormalizations $Z(C)$, $Z(\gamma_{(1)})$, and $Z(\gamma_{(2)})$ are not equal to each other or to Z_2. If we write

$$\begin{aligned}Z(C)&=1+\Lambda(C),\\ Z_2&=1+\Lambda_2,\end{aligned}\quad(14)$$

then we find, to second order, that

$$\begin{aligned}\tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,p',q)&=T_{(1)(2)}{}^{*\text{Compt}}(p,p',q)^{\text{finite}}\\ &\quad+[2\Lambda_2-\Lambda(\gamma_{(1)})-\Lambda(\gamma_{(2)})]\quad(15)\\ &\quad\times\left[\gamma_{(1)}\frac{1}{\gamma\cdot p+\gamma\cdot q}\gamma_{(2)}+\gamma_{(2)}\frac{1}{\gamma\cdot p-\gamma\cdot q'}\gamma_{(1)}\right],\\ \tilde{\Gamma}(C;p,p')&=\Gamma(C;p,p')^{\text{finite}}+[\Lambda_2-\Lambda(C)]C.\end{aligned}$$

[9] F. J. Gilman, Phys. Rev. 167, 1365 (1968).

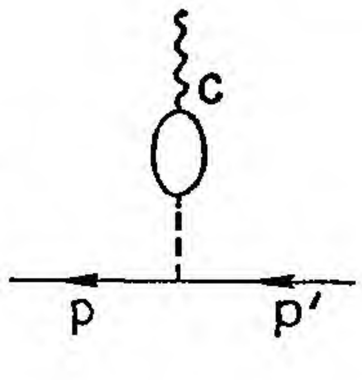


FIG. 4. Self-energy diagram which makes a second-order correction to the SU_3-singlet part of C.

Using the fact that finite quantities on the left- and right-hand sides of Eq. (7b) must match up, we see that

$$\Delta^{\text{Compt}}=[\Lambda_2+\Lambda(C)-\Lambda(\gamma_{(1)})-\Lambda(\gamma_{(2)})]C+\text{finite},\quad(16)$$

confirming that the $\ln\lambda^2$ dependence in Δ^{Compt} results from a mismatch between the multiplicative renormalization factors on the left- and right-hand sides of Eq. (7b). To check Eq. (16) directly, we note from Eq. (A10) that

$$\begin{aligned}\Lambda(C)C&=(g_r^2/32\pi^2)\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau C\gamma^\tau\boldsymbol{\gamma}\ln\lambda^2,\\ \Lambda_2\gamma_0&=(g_r^2/32\pi^2)\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_0\gamma^\tau\boldsymbol{\gamma}\ln\lambda^2,\end{aligned}\quad(17)$$

which allows us to rewrite the square bracket in Eq. (16) in the form

$$\begin{aligned}(g_r^2/32\pi^2)\ln\lambda^2\{&\gamma_{(1)}\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_0\gamma^\tau\boldsymbol{\gamma}\gamma_{(2)}-\gamma_{(2)}\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_0\gamma^\tau\boldsymbol{\gamma}\gamma_{(1)}\\ &+\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau[\gamma_{(1)}\gamma_0\gamma_{(2)}-\gamma_{(2)}\gamma_0\gamma_{(1)}]\gamma^\tau\boldsymbol{\gamma}\\ &-\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(1)}\gamma^\tau\boldsymbol{\gamma}\gamma_0\gamma_{(2)}+\gamma_{(2)}\gamma_0\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(1)}\gamma^\tau\boldsymbol{\gamma}\\ &-\gamma_{(1)}\gamma_0\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(2)}\gamma^\tau\boldsymbol{\gamma}+\tfrac{1}{2}\boldsymbol{\gamma}\gamma_\tau\gamma_{(2)}\gamma^\tau\boldsymbol{\gamma}\gamma_0\gamma_{(1)}\},\end{aligned}\quad(18)$$

with the four lines coming from Λ_2, $\Lambda(C)$, $-\Lambda(\gamma_{(1)})$, and $-\Lambda(\gamma_{(2)})$, respectively. A little algebra then shows that Eq. (18) is indeed identical with the $\ln\lambda^2$ part of Eq. (7c).

The presence of terms which diverge as $\ln|q_0|^2$ in Eq. (7c) indicates that, *in the general case, the Bjorken limit does not exist.* The fact that the $\ln|q_0|^2$ and $\ln\lambda^2$ terms occur in the combination $\ln(\lambda^2/|q_0|^2)$ means that, *to second order, the existence of the Bjorken limit is directly connected with the matching of renormalization factors on the left- and right-hand sides of Eq.* (7b): When the renormalization factors match, the Bjorken limit exists; when the factors do not match, the Bjorken limit diverges.[10] Unfortunately, we shall see that this simple connection does not hold in higher orders of perturbation theory.

To interpret the divergence of the Bjorken limit, we note that the renormalized T product can be written as[3]

$$\tilde{T}_{(1)(2)}(p,p',q)=\int_{-\infty}^{\infty}dq_0'\,\frac{\rho(p,\,p',\,\mathbf{q},\,q_0')}{q_0-q_0'},\quad(19)$$

[10] This was first noted by A. I. Vainshtein and B. L. Ioffe, Zh. Eksperim. i Teor. Fiz. Pis'ma v Redaktsiyu 6, 917 (1967) [Soviet Phys. JETP Letters 6, 341 (1967)]. These authors conjectured that when the renormalization factors match, the Bjorken limit and naive commutator agree. Our calculations show that this conjecture is invalid.

TABLE I. Cases involving vector (V) and axial-vector (A) octet currents with finite Bjorken limit in second order.

Model	Current $J_{(1)}$	Current $J_{(2)}$	Piece of current $\bar{\psi}C\psi$
Vector gluon	V or A	V or A	V or A
Scalar or pseudo-scalar gluon	V	V	V
	V	A	A
	A	V	A

where the spectral function ρ is defined by

$$\bar{u}(p)\rho(p,p',\mathbf{q},q_0)u(p') = (2\pi)^3 \sum_N \langle p | J_{(1)} | N\rangle\langle N | J_{(2)} | p'\rangle \delta^4(q+p-N) - (2\pi)^3 \sum_N \langle p | J_{(2)} | N\rangle\langle N | J_{(1)} | p'\rangle \delta^4(q+N-p'). \tag{20}$$

Provided that the spectral function does not oscillate an infinite number of times[11] (and it cannot have this kind of pathological behavior in perturbation theory), when the Bjorken limit of $\tilde{T}_{(1)(2)}(p,p',q)$ exists it is equal to the integral

$$\bar{u}(p)\int_{-\infty}^{\infty} dq_0'\, \rho(p,p',\mathbf{q},q_0')u(p') = (2\pi)^3 \sum_N \langle p | J_{(1)} | N\rangle\langle N | J_{(2)} | p'\rangle \delta^3(\mathbf{q}+\mathbf{p}-\mathbf{N}) - (2\pi)^3 \sum_N \langle p | J_{(2)} | N\rangle\langle N | J_{(1)} | p'\rangle \delta^3(\mathbf{q}+\mathbf{N}-\mathbf{p}'), \tag{21}$$

which is just the usual sum-over-intermediate-states definition of the commutator. Conversely, in the cases in which the Bjorken limit diverges logarithmically, the integral and sum in Eq. (21) must diverge logarithmically.

2. There are a number of interesting cases in which the renormalization factors do match, and hence the Bjorken limit exists in second order. We have enumerated in Table I all examples of this type in which all of the currents involved, $J_{(1)}$, $J_{(2)}$, and $\bar{\psi}C\psi$, are either vector or axial-vector octet currents. Specific formulas for Δ^{Compt} in the case when $J_{(1)}$ and $J_{(2)}$ are vector currents were given in Eqs. (10) and (11) above. (To obtain the corresponding formulas when $J_{(1)}$ and/or $J_{(2)}$ are axial-vector currents, in the vector-gluon case, one simply multiplies from the left or right by γ_5 according to the scheme shown in Table II.)

The remarkable result that emerges from these examples is that, *even when the Bjorken limit exists in second order, it does not agree with the naive commutator* (that is, Δ^{Compt} is finite but nonzero). According to our previous discussion, this means that the Bjorken limit agrees with the spectral function integral of Eq. (21), but the naive commutator does not. Most of the principal applications of the Bjorken limit technique for space-component–space-component commutators *assume* the identity of the Bjorken limit and naive commutator, and therefore, according to our results, break down in perturbation theory. Further details of this breakdown are given in Ref. 6.

3. From an inspection of Eq. (10) and Table II, we see that in the vector-gluon case, for all commutators involving vector and axial-vector currents, Δ^{Compt} vanishes when either $\mu=0$ or $\nu=0$. This result can be deduced directly from the Ward identity[10] satisfied by $\tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,p',q)$, which, in the case when $J_{(1)}$ and $J_{(2)}$ are both vector currents, states that

$$\tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,p',q)\,|_{\gamma(1)=q^\mu\gamma_\mu\lambda^a,\gamma(2)=\gamma_\nu\lambda^b} = \tilde{\Gamma}([\lambda^a,\lambda^b]\gamma_\nu;p,p'). \tag{22}$$

Multiplying by q_0^{-1} and taking the limit $q_0\to i\infty$ gives immediately

$$\lim_{q_0\to i\infty} \tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,p',q)\,|_{\gamma(1)=\gamma_0\lambda^a,\gamma(2)=\gamma_\nu\lambda^b} = q_0^{-1}\tilde{\Gamma}([\lambda^a,\lambda^b]\gamma_\nu;p,p')+O(q_0^{-2}\ln q_0), \tag{23}$$

confirming our explicit calculation. A similar derivation holds in the cases involving axial-vector currents, provided that the divergence of the axial-vector current is "soft,"[12] as it is in the vector-gluon case. We thus see that the breakdown of the Bjorken limit which we have found is consistent with the constraints imposed by Ward identities. Therefore all of the results of the Gell-Mann time-component algebra, which are derived directly from the Ward identities, remain valid.[13]

4. We turn next to the order g_r^4 result of Eq. (12), which gives the $VV\to V$ commutator in the scalar- and pseudoscalar-gluon models (the second line in Table I). We see that *even though the renormalization factors match, the Bjorken limit in this case diverges in fourth*

TABLE II. Substitutions to get axial-vector current results in the vector-gluon case.

Current $J_{(1)}$	Current $J_{(2)}$	Change in Eq. (10)
V	V	none
A	V	$\Delta^{\text{Compt}}\to -\gamma_5\Delta^{\text{Compt}}$
V	A	$\Delta^{\text{Compt}}\to\Delta^{\text{Compt}}\gamma_5$
A	A	$\Delta^{\text{Compt}}\to -\gamma_5\Delta^{\text{Compt}}\gamma_5$

[11] This pathological case is discussed in detail by R. Brandt and J. Sucher, Phys. Rev. Letters 20, 1131 (1968).

[12] See Ref. 2, pp. 257–260, for a discussion of soft divergences.

[13] There is one exception to this statement, which arises when Ward identity anomalies are present. See S. L. Adler, Phys. Rev. 177, 2426 (1969); J. S. Bell and R. Jackiw, Nuovo Cimento 60, 47 (1969); R. Jackiw and K. Johnson, Phys. Rev. 182, 1459 (1969); S. L. Adler and W. A. Bardeen, *ibid.* 182, 1517 (1969).

order. We note, however, that the divergence behaves as $g_r^4 \ln|q_0|^2$, whereas in fourth order, terms behaving like $g_r^4(\ln|q_0|^2)^2$ could in principle be present. On the basis of this behavior and our second-order results, we make the following *conjecture: When the renormalization factors needed to make $T_{(1)(2)}^*(p, p', q)$ and $\Gamma(C; p, p')$ finite are the same, the Bjorken limit in order 2n of perturbation theory contains no terms $g_r^{2n}(\ln|q_0|^2)^n$, but begins in general with terms $g_r^{2n}(\ln|q_0|^2)^{n-1}$.*

We have only calculated results for the scalar- and pseudoscalar-gluon models because these models have the simple property that, when the unitarity method of Appendix B is used, each individual intermediate state makes a contribution behaving at worst as $g_r^4 \ln|q_0|^2$. The situation in the vector-gluon model is more complicated, since here the individual intermediate states contain terms behaving as $g_r^4(\ln|q_0|^2)^2$, as well as terms $g_r^4 \ln|q_0|^2$. If our conjecture is correct, the $g_r^4(\ln|q_0|^2)^2$ terms from the various intermediate states in the vector-gluon case must add up to zero. We have not checked whether this happens; it would clearly be worth doing.

5. As mentioned in Sec. I, one can try to save asymptotic sum rules by postulating that nonperturbative effects conspire to make asymptotic sum rules valid when all orders of perturbation theory are summed. A simple way that this could happen would be if our order g_r^2 terms in Δ^{Compt} were the lowest-order terms in an expression

$$A \exp[-Bg_r^2 \ln|q_0|^2], \qquad B>0 \tag{24}$$

which damps to zero as $q_0 \to i\infty$. However, examination of our fourth-order result in Eq. (12) shows that exponentiation gives

$$\frac{g_r^2}{16\pi^2}+7\left(\frac{g_r^2}{16\pi^2}\right)^2 \ln|q_0|^2 \approx \frac{g_r^2}{16\pi^2} \times\exp\left(7\frac{g_r^2}{16\pi^2}\ln|q_0|^2\right)+O(g_r^6), \tag{25}$$

which blows up exponentially rather than damping. In other words, the simple damping mechanism of Eq. (24) cannot be correct, although our fourth-order calculation obviously cannot rule out more complicated damping mechanisms.

6. In Eqs. (A14) and (A15) of Appendix A, we indicate that when the Bjorken limit $q_0 \to i\infty$ is taken *before* letting the regulator mass λ go to infinity, one obtains just the naive commutator. Thus, it is tempting to try to "save" asymptotic sum rules by prescribing that, instead of using renormalized perturbation theory (limit $\lambda \to \infty$ taken first), one should always work with the unrenormalized quantities, with λ very large but finite.[14] We will now argue, however, that this is a spurious resolution of the difficulty. Let us consider the sum rule, derived in Appendix B, connecting the $[\lambda^a, \lambda^b]$ term in Eq. (11) with an integral over the longitudinal current-nucleon cross section $L^-(q^2, \omega)$, with $\omega = -q^2/p\cdot q$. In the renormalized ($\lambda \to \infty$) theory, where there is Bjorken-limit breakdown, we find *to second order* that

$$\lim_{q_0\to i\infty} \tilde{T}_{(1)(2)}^{*\text{Compt}} = q_0^{-1}\tfrac{1}{2}[\lambda^a, \lambda^b] \times[\gamma_\mu\gamma_0\gamma_\nu+\gamma_\nu\gamma_0\gamma_\mu+2(g_{\mu\nu}-g_{\mu 0}g_{\nu 0})\gamma_0 f] + (\text{term symmetric in } a, b)+O(q_0^{-2}\ln q_0), \tag{26}$$

$$f=\lim_{q^2\to-\infty} 2\int_0^2 d\omega\, L_{n+g}^-(q^2, \omega), \tag{27}$$

where the subscript $n+g$ is a reminder that to second order we need only retain the single neutron plus gluon intermediate-state contribution in calculating L^-. Our explicit calculation shows that

$$f=g_r^2/16\pi^2,$$
$$\lim_{q^2\to-\infty} L_{n+g}^-(q^2, \omega)=(g_r^2/64\pi^2)\omega, \tag{28}$$

in agreement with Eq. (27). As we noted in Ref. 6, Eq. (27) indicates that the breakdown of the Bjorken limit in Eq. (26) is essentially the same phenomenon as the breakdown of the Callan-Gross relation,[15] which states that

$$\lim_{q^2\to-\infty} L_{n+g}^-(q^2, \omega)=0. \tag{29}$$

Let us now consider the analogs of Eqs. (26)–(29) in the regulated (λ-finite) theory. Since, in order g_r^2, matrix elements are always *linear* in the gluon propagator, to obtain the regulated matrix element in order g_r^2 we simply subtract from the renormalized matrix element the corresponding expression with the gluon mass μ^2 replaced by the regulator mass λ^2. Since f in Eq. (28) is independent of μ^2, we find that Eqs. (26)–(28) become

$$\lim_{q_0\to i\infty} T_{(1)(2)}^{*\text{Compt}} = q_0^{-1}\tfrac{1}{2}[\lambda^a, \lambda^b](\gamma_\mu\gamma_0\gamma_\nu+\gamma_\nu\gamma_0\gamma_\mu) + (\text{term symmetric in } a, b)+O(q_0^{-2}\ln q_0), \tag{30}$$

$$0=\lim_{q^2\to-\infty} 2\int_0^2 d\omega\, L_{\text{tot}}^-(q^2, \omega);$$
$$\lim_{q^2\to-\infty} L_{\text{tot}}^-(q^2, \omega)=0, \tag{31}$$

with

$$L_{\text{tot}}^-(q^2, \omega)=L_{n+g}^-(q^2, \omega)-L_{n+g}^-(q^2, \omega)\,|_{\mu^2\to\lambda^2}. \tag{32}$$

As expected, in the regulated theory the Bjorken limit is normal and the Callan-Gross relation is satisfied. However, a disturbing problem arises when we examine in detail exactly *how* the Bjorken limit is satisfied. Let us suppose that the regulator mass λ is much larger than

[14] This point of view has been advocated by C. R. Hagen, Phys. Rev. 188, 2416 (1969).

[15] C. G. Callan and D. J. Gross, Phys. Rev. Letters 22, 156 (1969).

the fermion mass m and the gluon mass μ,

$$\lambda^2 \gg m^2, \quad \lambda^2 \gg \mu^2, \tag{33}$$

and let us consider, for fixed ω, two ranges of values of $-q^2$,

$$\begin{aligned} &\text{range 1:} \quad \mu^2, m^2 \ll -q^2 < \xi/(2\omega^{-1}-1), \\ &\text{range 2:} \quad \xi/(2\omega^{-1}-1) \le -q^2, \quad \xi = (\lambda+m)^2 - m^2. \end{aligned} \tag{34}$$

The dividing point between the two ranges is just the threshold for regulator particle production. For $-q^2 < \xi/(2\omega^{-1}-1)$, we have $(q+p)^2 < (\lambda+m)^2$, and regulator particle production is forbidden. Thus, in range 1, the second term in Eq. (32) vanishes,

$$L_{n+g}^{-}(q^2, \omega)\,|_{\mu^2\to\lambda^2} = 0, \tag{35}$$

while the first term has its asymptotic value

$$L_{n+g}^{-}(q^2, \omega) \approx (g_r^2/64\pi^2)\omega, \tag{36}$$

and the Callan-Gross limit is *not* satisfied. In range 2, we have $(q+p)^2 \ge (\lambda+m)^2$, and regulator production is allowed; for $-q^2 \gg \lambda^2/(2\omega^{-1}-1)$, the second term in Eq. (32) attains the same asymptotic value as Eq. (36), and the Callan-Gross limit is satisfied. Thus, we see that *in the regulator theory, the Callan-Gross limit is satisfied only in a region in which $-q^2$ is big on a scale determined by λ^2, and then only by virtue of the unphysical, negative contribution of regulator production to the total longitudinal cross section.* We conclude that the regulator theory does not afford a satisfactory resolution of the breakdown of the Bjorken limit in perturbation theory.

ACKNOWLEDGMENTS

We wish to thank W. A. Bardeen, S. D. Drell, D. J. Gross, and S. B. Treiman for helpful discussions. One of us (S.L.A.) wishes to thank Professor A. Pais for the hospitality of the Rockefeller University, and the other (W.-K.T.) acknowledges the hospitality of SLAC, where portions of this work were done.

APPENDIX A: CALCULATION OF Δ^{Compt}

In this appendix we outline the calculation leading to Eq. (7) in the text. We recall that $\tilde{T}_{(1)(2)}{}^{*\text{Compt}}$ is defined as the contribution to the current-fermion scattering amplitude of the diagrams shown in Fig. 1, consisting of the lowest-order current-fermion diagrams and the second-order diagrams obtained from the lowest-order ones by insertion of a single virtual gluon. We may write

$$\begin{aligned} \tilde{T}_{(1)(2)}{}^{*\text{Compt}}(p,p',q) = {}& \tilde{\Gamma}(\gamma_{(1)};p,p+q)\tilde{S}(p+q) \\ &\times\tilde{\Gamma}(\gamma_{(2)};p+q,p') + \tilde{\Gamma}(\gamma_{(2)};p,p-q')\tilde{S}(p-q') \\ &\times\tilde{\Gamma}(\gamma_{(1)};p-q',p') + B_{(1)(2)}(p,p',q), \end{aligned} \tag{A1}$$

where $\tilde{S}$ and $\tilde{\Gamma}$ are the renormalized propagator and vertex functions and where $B_{(1)(2)}$ denotes the sum of the two box diagrams on the fifth line of Fig. 1. We shall calculate $\tilde{T}_{(1)(2)}{}^{*\text{Compt}}$ in the limit $q_0 \to i\infty$ and isolate the coefficient of the q_0^{-1} term. This is to be compared with the matrix element of the naive commutator between fermion states, given by $\tilde{\Gamma}(C;p,p')$.

The renormalized vertex function $\tilde{\Gamma}$ for the current $\bar{\psi}\gamma_{(1)}\psi$ is given, to second order in g, by

$$\tilde{\Gamma}(\gamma_{(1)};p,p') = Z_2\Gamma(\gamma_{(1)};p,p') = Z_2\gamma_{(1)} + \Lambda(\gamma_{(1)};p,p'), \tag{A2}$$

with Z_2 the fermion wave-function renormalization and with $\Lambda(\gamma_{(1)};p,p')$ the usual unrenormalized second-order vertex part (arising from diagrams on the second and fourth lines of Fig. 1). Note that $\tilde{\Gamma}$ is obtained by multiplying the unrenormalized vertex function by the wave-function renormalization, with *no further subtractions or rescaling.* The renormalized propagator is given, to second order in g_r, by the usual expression[16]

$$\tilde{S}(p)^{-1} = Z_2S(p)^{-1} = Z_2(\gamma\cdot p - m_0) - \Sigma(p), \tag{A3}$$

with $\Sigma(p)$ the unrenormalized proper fermion self-energy part (arising from the diagrams on the third line of Fig. 1) and with $m_0 = m + \delta m$ the fermion bare mass. Denoting the lowest-order current-fermion amplitude by $T_{(1)(2)}{}^{\text{Born}}$, we see that the first two lines on the right-hand side of Eq (A1) may be rewritten as

$$\begin{aligned} Z_2T_{(1)(2)}{}^{\text{Born}} + [&\Lambda(\gamma_{(1)};p,p+q)(\gamma\cdot p+\gamma\cdot q-m)^{-1}\gamma_{(2)} \\ &+\gamma_{(1)}(\gamma\cdot p+\gamma\cdot q-m)^{-1}(\delta m+\Sigma(p+q)) \\ &\times(\gamma\cdot p+\gamma\cdot q-m)^{-1}\gamma_{(2)}+\gamma_{(1)}(\gamma\cdot p+\gamma\cdot q-m)^{-1} \\ &\times\Lambda(\gamma_{(2)};p+q,p') + ((1)\leftrightarrow(2), q\leftrightarrow -q')]. \end{aligned} \tag{A4}$$

According to Eq. (A2), the matrix element of the naive commutator is

$$Z_2C + \Lambda(C;p,p'). \tag{A5}$$

It is easy to see that, as $q_0 \to i\infty$, the q_0^{-1} term of $Z_2T_{(1)(2)}{}^{\text{Born}}$ is precisely Z_2C. Our task is therefore reduced to comparing the q_0^{-1} term of

$$\begin{aligned} [&\Lambda(\gamma_{(1)};p,p+q)(\gamma\cdot p+\gamma\cdot q-m)^{-1}\gamma_{(2)}+\cdots \\ &+((1)\leftrightarrow(2), q\leftrightarrow -q')] + B_{(1)(2)}(p,p',q) \end{aligned} \tag{A6}$$

with $\Lambda(C;p,p')$.

The unrenormalized self-energy and vertex parts Σ and Λ are calculated by the usual technique of introducing a meson regulator of mass λ, giving

$$\Sigma(p) = \frac{-g_r^2}{16\pi^2}\int_0^1 dx\, \gamma(x\gamma\cdot p+m)\ln\gamma\left[\frac{x(1-x)(-p^2+m^2)+x\lambda^2+(1-x)^2m^2}{x(1-x)(-p^2+m^2)+x\mu^2+(1-x)^2m^2}\right] \tag{A7}$$

[16] In this equation, $S(p)$ denotes the unrenormalized propagator in the presence of all interactions.

and

$$\Lambda(\gamma_{(1)};p,p')=\frac{-g_r^2}{16\pi^2}\int_0^1 dx\int_0^{1-x}dy\bigg\{\tfrac{1}{2}\gamma\gamma_r\gamma_{(1)}\gamma^r\gamma\ln\bigg[\frac{z\lambda^2-C(p,p',x,y)}{z\mu^2-C(p,p',x,y)}\bigg]$$
$$+\gamma[(1-x)\gamma\cdot p-y\gamma\cdot p'+m]\gamma_{(1)}[(1-y)\gamma\cdot p'-x\gamma\cdot p+m]\gamma\bigg[\frac{1}{C(p,p',x,y)-z\mu^2}-\frac{1}{C(p,p',x,y)-z\lambda^2}\bigg]\bigg\},\quad \text{(A8)}$$

where

$$z=1-x-y,$$
$$C(p,p',x,y)=x(1-x)p^2+y(1-y)p'^2-2xyp\cdot p'-(x+y)m^2. \quad \text{(A9)}$$

In order to obtain the renormalized propagator and vertex from Eqs. (A2) and (A3), we must calculate the $\lambda\to\infty$ limit of the unrenormalized self-energy and vertex parts, dropping all terms which vanish in this limit but retaining powers of $\ln\lambda$. [Note that the $\lambda\to\infty$ limits of Σ and Λ are *not* the same as the *renormalized* self-energy and vertex parts $\tilde{\Sigma}$ and $\tilde{\Lambda}$, which are defined, in the $\lambda\to\infty$ limit, by $Z_2(\gamma\cdot p-m_0)-\Sigma(p)=\gamma\cdot p-m-\tilde{\Sigma}(p)$, $Z_2\gamma_{(1)}+\Lambda(\gamma_{(1)};p,p')=\gamma_{(1)}+\tilde{\Lambda}(\gamma_{(1)};p,p')$.] Taking the $\lambda\to\infty$ limit in Eqs. (A7)-(A9), we find

$$\lim_{\lambda\to\infty}\Sigma(p)=\frac{-g_r^2}{16\pi^2}\int_0^1 dx\,\gamma(x\gamma\cdot p+m)\gamma\ln\bigg[\frac{x\lambda^2}{x(1-x)(-p^2+m^2)+x\mu^2+(1-x)^2m^2}\bigg],$$

$$\lim_{\lambda\to\infty}\Lambda(\gamma_{(1)};p,p')=\frac{-g_r^2}{16\pi^2}\int_0^1 dx\int_0^{1-x}dy\bigg\{\tfrac{1}{2}\gamma\gamma_r\gamma_{(1)}\gamma^r\gamma\ln\bigg[\frac{z\lambda^2}{z\mu^2-C(p,p',x,y)}\bigg]$$
$$+\gamma[(1-x)\gamma\cdot p-y\gamma\cdot p'+m]\gamma_{(1)}[(1-y)\gamma\cdot p'-x\gamma\cdot p+m]\gamma\frac{1}{C(p,p',x,y)-z\mu^2}\bigg\}.\quad \text{(A10)}$$

Finally, taking infinite-momentum limits of these expressions, we find

$$\lim_{p_0\to i\infty}\lim_{\lambda\to\infty}\Sigma(p)=(-g_r^2/16\pi^2)\gamma\gamma_0\gamma p_0[\tfrac{1}{2}\ln(\lambda^2/|p_0|^2)+\tfrac{3}{4}]+O(\ln p_0),$$
$$\lim_{p_0\to i\infty}\lim_{\lambda\to\infty}\Lambda(\gamma_{(1)};p,p')=(g_r^2/16\pi^2)\{-\tfrac{1}{4}\gamma\gamma_r\gamma_{(1)}\gamma^r\gamma[\ln(\lambda^2/|p_0|^2)+\tfrac{1}{2}]+\tfrac{1}{2}\gamma\gamma_0\gamma_{(1)}\gamma_0\gamma+O(\ln p_0/p_0)\},\quad \text{(A11)}$$
$$\lim_{p_0'\to i\infty}\lim_{\lambda\to\infty}\Lambda(\gamma_{(1)};p,p')=(g_r^2/16\pi^2)\{-\tfrac{1}{4}\gamma\gamma_r\gamma_{(1)}\gamma^r\gamma[\ln(\lambda^2/|p_0'|^2)+\tfrac{1}{2}]+\tfrac{1}{2}\gamma\gamma_0\gamma_{(1)}\gamma_0\gamma+O(\ln p_0'/p_0')\}.$$

The box diagram is convergent even without regularization. In the regulator theory, $B_{(1)(2)}$ is the difference of two terms calculated with meson masses μ and λ, respectively, but the term with mass λ does not contribute in the limit $\lambda\to\infty$. [The situation is similar to the second term on the right-hand side of Eq. (A8).] A little care must be exercised in computing the Bjorken limit of $B_{(1)(2)}$. The reason is that, because of infrared singularities, the limit $q_0\to i\infty$ cannot be naively taken under the integrals over the Feynman parameters.[17] A detailed study yields

$$\lim_{q_0\to i\infty}\lim_{\lambda\to\infty}B_{(1)(2)}(p,p',q)=q_0^{-1}\Lambda(C;p,p')$$
$$+(g_r^2/16\pi^2)q_0^{-1}\{\tfrac{1}{4}\gamma\gamma_r\gamma_{(1)}\gamma_0\gamma_{(2)}\gamma^r\gamma\ln(\lambda^2/|q_0|^2)$$
$$+\tfrac{1}{8}\gamma[\gamma_r\gamma_{(1)}\gamma^r\gamma_{(2)}\gamma_0+\gamma_0\gamma_{(1)}\gamma_r\gamma_{(2)}\gamma^r]\gamma$$
$$-\tfrac{1}{4}\gamma\gamma_0\gamma_{(1)}\gamma_0\gamma_{(2)}\gamma_0\gamma-[(1)\leftrightarrow(2)]\}+O(\ln q_0/q_0^2).\quad \text{(A12)}$$

[17] The problem which one encounters here can be illustrated by a simple example. Consider the integral $\int_0^1 dx\,(Ax^2q^2+B)/(xq^2+C)^2$, with A, B, and C constants. In the limit $q^2\to-\infty$, both terms in the numerator behave as $(q^2)^{-1}$, although at first glance one might expect the second term to behave like $(q^2)^{-2}$ and to be negligible compared to the first.

Note that, according to Eqs. (3) and (A11), the $\ln\lambda^2$ dependence of $\Lambda(C;p,p')$ *precisely cancels* the $\ln\lambda^2$ in the curly bracket in Eq. (A12), as required by the absence of $\ln\lambda^2$ dependence on the left-hand side. Substituting Eqs. (A11) and (A12) into Eq. (A6), we obtain, finally,

$$\lim_{q_0\to i\infty}\tilde{T}_{(1)(2)}{}^{*\mathrm{Compt}}(p,p',q)=(1/q_0)$$
$$\times[\tilde{\Gamma}(C;p,p')+\Delta^{\mathrm{Compt}}]+O(q_0^{-2}\ln q_0),\quad \text{(A13)}$$

with Δ^{Compt} as given by Eq. (7c) of the text.

To conclude this appendix we remark that if, starting from the regulated quantities of Eqs. (A7) and (A8) and the regulated box-diagram part $B_{(1)(2)}$, one took the Bjorken limit $q_0\to i\infty$ *before* letting the regulator mass λ go to infinity, one would obtain

$$\lim_{p_0\to i\infty}\Sigma(p)=O(1/p_0),$$
$$\lim_{p_0\to i\infty}\Lambda(\gamma_{(1)};p,p')=O(1/p_0),$$
$$\lim_{p_0\to i\infty}\Lambda(\gamma_{(1)};p,p')=O(1/p_0'),\quad \text{(A14)}$$
$$\lim_{q_0\to i\infty}B_{(1)(2)}(p,p',q)=(1/q_0)\Lambda(C;p,p')+O(1/q_0^2).$$

TABLE III. Regions of phase space where denominators in Eq. (2.19) vanish as $\mu^2 \to 0$. n^s, g_1^s, $\cdots$ denote the spatial components ($s=1,2,3$) of n, g_1, $\cdots$.

	Phase-space region	Denominators which vanish
(1)	$n^s=0$	$(n+g_2)^2$, $(n+g_1)^2$
(2)	$g_1^s=0$	$(n+g_1)^2$, $(p-g_1)^2$
(3)	$g_2^s=0$	$(n+g_2)^2$, $(p-g_2)^2$
(4)	$g_1^s \parallel p^s$	$(p-g_1)^2$
(5)	$g_2^s \parallel p^s$	$(p-g_2)^2$
(6)	$g_1^s \parallel p^s$ and $g_2^s \parallel p^s$	$(p-g_1)^2$, $(p-g_2)^2$, $(p-g_1-g_2)^2$
(7)	$g_1^s \parallel n^s$	$(n+g_1)^2$
(8)	$g_2^s \parallel n^s$	$(n+g_2)^2$

As a consequence, one finds

$$\lim_{q_0 \to i\infty} T_{(1)(2)}{}^{*\mathrm{Compt}} = (1/q_0)\Gamma(C;p,p') + O(1/q_0^2); \quad \text{(A15)}$$

that is, the Bjorken limit in the case of finite regulator mass agrees with the naive commutator. This result is expected for the regulator theory since the anomalous term Δ^{Compt} is independent of the gluon mass and is canceled by exactly the same term (with opposite sign) which must be present when the regulator mass is kept finite.

APPENDIX B: FOURTH-ORDER CALCULATION

In this appendix we consider an extension of our previous results to order g_r^4. Unfortunately, repeating the general calculation of Appendix A in the next order of perturbation theory would require a prohibitive amount of work, and therefore will not be attempted. Rather, we will content ourselves with the calculation of one special case, which is made tractable by a combination of tricks. The special case is the SU_3-antisymmetric piece of the vector-vector commutator, in the scalar- and pseudoscalar-gluon models. There are two further restrictions. We consider *only* the leading logarithmic behavior in the Bjorken limit, and we limit ourselves to the part of the commutator which, like Δ^{Compt}, is *independent* of the three-momenta $\mathbf{q}$, $\mathbf{q}'$, $\mathbf{p}$, and $\mathbf{p}'$ and of the fermion mass m. This second restriction means that we can set $\mathbf{q}=\mathbf{q}'=\mathbf{p}=\mathbf{p}'=\mathbf{0}$ at the outset, so that we are dealing with the forward Compton scattering amplitude, and that we can take the limit $m\to 0$ wherever $\ln m$ divergences do not appear. (We will verify that there are no $\ln m$ factors in the leading $\ln|q_0|^2$ term.) The restrictions allow us to employ the following two tricks, which make the calculation tractable: (i) We exploit a connection, provided by unitarity, between the Bjorken limit of the forward Compton amplitude and current-fermion cross sections. This connection becomes especially simple in the $m\to 0$ limit. (ii) For dimensional reasons, $\ln|q_0|^2$ terms in the current-fermion cross section (at $m=0$) must be accompanied by $-\ln\mu^2$ terms, where μ is the gluon mass, so we can study the large-$|q_0|^2$ behavior by studying the small-μ^2 singularities. The latter arise from readily identifiable regions of phase space, and are much more easily evaluated than the complete current-fermion cross section itself.

We begin by reviewing the unitarity connection[6,9] between the current-fermion cross sections and the forward Compton amplitude. Since we are only interested in the commutator of two vector currents, we set $\gamma_{(1)}=\gamma_\mu\lambda^a$, $\gamma_{(2)}=\gamma_\nu\lambda^b$, $J_{(1)}=J_\mu{}^a$, and $J_{(2)}=J_\nu{}^b$. It will further be convenient to restrict a and b to lie in the isospin SU_2 subspace of SU_3 (a, $b=1$, 2); this has no effect on the part of the commutator antisymmetric in a and b, and has the virtue of making the charge structure of our problem identical to the familiar case of pion-nucleon scattering. Denoting $\omega=-q^2/p\cdot q$, we may write for the spin-averaged, forward-scattering current-"proton" amplitude,[18]

$$\tfrac{1}{4}\mathrm{Tr}\left[\left(\frac{\gamma\cdot p+m}{2m}\right)\tilde{T}_{(1)(2)}{}^*(p,p,q)\right] = \text{polynomial} + \left(\frac{-i}{4}\right)\sum_{\mathrm{spin}(p)}\int d^4x\, e^{iq\cdot x}\langle p|T(J_\mu{}^a(x)J_\nu{}^b(0))|p\rangle$$

$$= \tfrac{1}{4}\mathrm{Tr}\left[\left(\frac{\gamma\cdot p+m}{2m}\right)\left(\gamma_\mu\lambda^a\frac{1}{\gamma\cdot p+\gamma\cdot q-m}\gamma_\nu\lambda^b + \gamma_\nu\lambda^b\frac{1}{\gamma\cdot p-\gamma\cdot q-m}\gamma_\mu\lambda^a\right)\right]$$

$$+ T_1{}^{ab}(q^2,\omega)\left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right) + T_2{}^{ab}(q^2,\omega)\left(p_\mu - \frac{p\cdot q}{q^2}q_\mu\right)\left(p_\nu - \frac{p\cdot q}{q^2}q_\nu\right). \quad \text{(B1)}$$

On the third and fourth lines, we have explicitly separated off the Born approximation and made use of the vector Ward identities for $\tilde{T}_{(1)(2)}{}^*(p,p,q)$, which imply that the non-Born part is divergenceless. The isospin structure of the non-Born amplitudes may be written in the form

$$T_{1,2}{}^{ab}(q^2,\omega) = T_{1,2}{}^{(+)}(q^2,\omega)\tfrac{1}{2}\{\lambda^a,\lambda^b\} + T_{1,2}{}^{(-)}(q^2,\omega)\tfrac{1}{2}[\lambda^a,\lambda^b]. \quad \text{(B2)}$$

The standard forward dispersion relations analysis for pion-nucleon scattering[19] may now be taken over to show

[18] Here "proton" means the p-type quark, and similarly "neutron" means the n-type quark. The actual matrix element is obtained by sandwiching Eq. (B1) between "proton" isospinors.

[19] G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. 106, 1337 (1957).

that the amplitudes $T_{1,2}^{(\pm)}$ satisfy the following dispersion relations:

$$T_1^{(+)}(q^2,\omega)=T_1^{(+)}(q^2,\infty)-\int_0^2 d\omega'[W_1^-(q^2,\omega')+W_1^+(q^2,\omega')][(\omega'-\omega)^{-1}+(\omega'+\omega)^{-1}],$$

$$T_1^{(-)}(q^2,\omega)=-\int_0^2 d\omega'\,[W_1^-(q^2,\omega')-W_1^+(q^2,\omega')][(\omega'-\omega)^{-1}-(\omega'+\omega)^{-1}],$$

$$T_2^{(+)}(q^2,\omega)=-\omega\int_0^2\frac{d\omega'}{\omega'}[W_2^-(q^2,\omega')+W_2^+(q^2,\omega')][(\omega'-\omega)^{-1}-(\omega'+\omega)^{-1}], \tag{B3}$$

$$T_2^{(-)}(q^2,\omega)=-\omega\int_0^2\frac{d\omega'}{\omega'}[W_2^-(q^2,\omega')-W_2^+(q^2,\omega')][(\omega'-\omega)^{-1}+(\omega'+\omega)^{-1}],$$

with absorptive parts given by

$$-(2\pi)^3\tfrac{1}{4}\sum_{\mathrm{spin}(p)}\sum_N\langle p\,|\,2^{-1/2}(J_\mu{}^1\mp iJ_\mu{}^2)\,|\,N\rangle\langle N\,|\,2^{-1/2}(J_\nu{}^1\pm iJ_\nu{}^2)\,|\,p\rangle\delta^4(p+q-N)$$

$$=2W_1^\pm(q^2,\omega)\left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right)+2W_2^\pm(q^2,\omega)\left(p_\mu-\frac{p\cdot q}{q^2}q_\mu\right)\left(p_\nu-\frac{p\cdot q}{q^2}q_\nu\right). \tag{B4}$$

In writing Eq. (B3), we have assumed one subtraction each for $T_1^{(\pm)}$ [the subtraction constant $T_1^{(-)}(q^2,\infty)$ vanishes by crossing symmetry] and no subtraction for $T_2^{(\pm)}$. To second order in g_r^2 we have explicitly checked the validity of these assumptions. Since the asymptotic behavior as $\omega\to 0$ of higher orders of perturbation theory will differ from second order only by powers of $\ln\omega$, and not by powers of ω, we expect these assumptions to be true to arbitrary order, and in particular, to order g_r^4.

Let us now set $\mathbf{q}=\mathbf{p}=0$, $\mu=\nu=1$ in Eq. (B1) and take the limit $q_0\to i\infty$. Using Eqs. (B2) and (B3), we find that the right-hand side of Eq. (B1) becomes

$$q_0^{-1}\tfrac{1}{2}[\lambda^a,\lambda^b]$$
$$\times\left\{1-\lim_{q^2\to-\infty}2m\int_0^2 d\omega'\,[W_1^-(q^2,\omega')-W_1^+(q^2,\omega')]\right\}$$
$$+(\text{term symmetric in } a,b)+O(q_0^{-2}\ln q_0). \tag{B5}$$

We know that the Bjorken limit of $\tilde{T}_{(1)(2)}{}^*(p,p,q)$ must have the general form

$$\lim_{q_0\to i\infty;\,\mathbf{q}=\mathbf{p}=0}\tilde{T}_{(1)(2)}{}^*=q_0^{-1}\tfrac{1}{2}[\lambda^a,\lambda^b]$$
$$\times[\gamma_\mu\gamma_0\gamma_\nu+\gamma_\nu\gamma_0\gamma_\mu+2(g_{\mu\nu}-g_{\mu0}g_{\nu0})\gamma_0 f(q^2/\mu^2,m^2/\mu^2)]$$
$$+(\text{term symmetric in } a,b)+O(q_0^{-2}\ln q_0), \tag{B6}$$

with f the difference between the Bjorken limit and the naive commutator. Setting $\mu=\nu=1$ in Eq. (B6), substituting for the left-hand side of Eq. (B1), and comparing with Eq. (B5), we get a sum rule for f,

$$f(q^2/\mu^2,m^2/\mu^2)=\lim_{q^2\to-\infty}2m\int_0^2 d\omega'$$
$$\times[W_1^-(q^2,\omega')-W_1^+(q^2,\omega')]. \tag{B7}$$

Equation (B7) can be rewritten in a more useful form by recalling that the usual fixed-q^2 sum rule, following from the Gell-Mann time-component algebra, is

$$0=\int_0^2\frac{d\omega'}{\omega'^2}[W_2^-(q^2,\omega')-W_2^+(q^2,\omega')], \tag{B8}$$

and is valid to all orders in perturbation theory in our models. Multiplying Eq. (B8) by $2mq^2$ and adding to Eq. (B7), we get the modified sum rule

$$f(q^2/\mu^2,m^2/\mu^2)=\lim_{q^2\to-\infty}2\int_0^2 d\omega'$$
$$\times[L^-(q^2,\omega')-L^+(q^2,\omega')], \tag{B9}$$

with

$$L^\mp(q^2,\omega)=2m[W_1^\mp(q^2,\omega)+(q^2/\omega^2)W_2^\mp(q^2,\omega)], \tag{B10}$$

the total longitudinal cross section for current-fermion scattering. The great virtue of Eq. (B9) is that, in the limit $m\to 0$, the longitudinal cross sections are given by the simple formula[20]

$$L^\mp=-(m\omega^2/q^2)(2\pi)^3$$
$$\times\tfrac{1}{4}\sum_{\mathrm{spin}(p)}\sum_N|\langle p\,|\,p^r\tfrac{1}{2}(J_r{}^1\pm iJ_r{}^2)\,|\,N\rangle|^2\delta^4(p+q-N), \tag{B11}$$

as may be readily verified by comparison of Eqs. (B11) and (B4). We will see that the factor p^r in Eq. (B11) enormously simplifies the subsequent calculation.

We are now ready to proceed with the calculation of f to order g_r^4. Before doing this, however, let us illustrate the procedure and check the arithmetic done so far by using Eqs. (B9) and (B11) to recalculate the order g_r^2 result contained in Eq. (11) of the text. To second order, the intermediate states which may contribute are the single "neutron,"[18] $N=n$, and the "neutron" plus gluon, $N=n+g$ (Fig. 5). Neither of these contributes to L^+, and the single "neutron" contribution to L^- vanishes to order g_r^2, because the zeroth-order

[20] We wish to thank D. J. Gross for pointing this out to us.

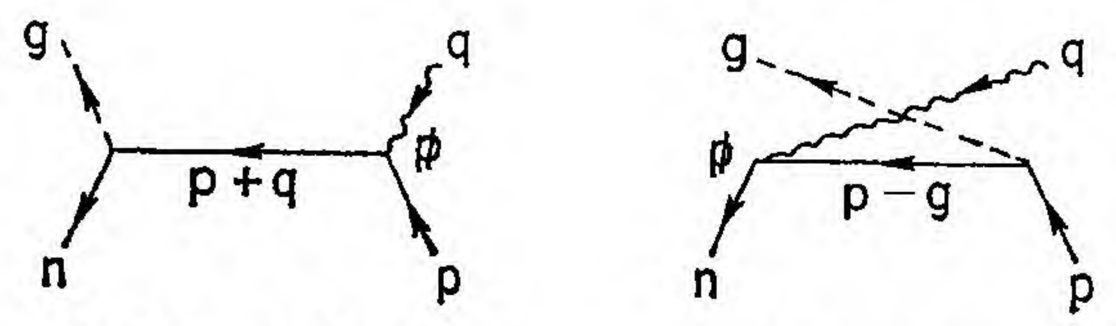


FIG. 5. Diagrams of order g_r contributing to the "neutron" plus gluon intermediate state.

part of $\langle p\,|\,p^{\tau}(J_r{}^1+iJ_r{}^2)\,|\,n\rangle$ is proportional to $\bar{u}(p)\gamma\cdot pu(n)=0$. So we have

$$L^+=0,$$

$$L^-=\frac{-m\omega^2}{q^2}(2\pi)^3\tfrac{1}{4}\sum_{\text{spin}(p)}\sum_{\text{spin}(n)}\int\frac{d^3n}{(2\pi)^3}$$

$$\times\frac{m}{n^0}\int\frac{d^3g}{(2\pi)^3}\frac{1}{2g^0}\delta^4(p+q-n-g)\,|\,\mathfrak{M}\,|^2, \qquad \text{(B12)}$$

$$\mathfrak{M}=g_r\bar{u}(n)\left(\frac{1}{\gamma\cdot p+\gamma\cdot q}\gamma\cdot p+\gamma\cdot p\,\frac{1}{\gamma\cdot p-\gamma\cdot g}\right)u(p),$$

with the factors $\gamma\cdot p$ in $\mathfrak{M}$ a result of the factor p^{τ} multiplying the current in Eq. (B11).[21] The factor $(\gamma\cdot p+\gamma\cdot q)^{-1}=(\gamma\cdot p+\gamma\cdot q)/[p\cdot q(2-\omega)]$ in the first term in $\mathfrak{M}$ would, if it survived, lead to a divergence in Eq. (B9) at the end point $\omega=2$, but it vanishes on account of the $\gamma\cdot p$ in the numerator. The second term in $\mathfrak{M}$ is also simplified by the presence of $\gamma\cdot p$, since it can be written as

$$g_r\bar{u}(n)[-2p\cdot g/(\mu^2-2p\cdot g)]u(p),$$

which approaches the finite quantity $g_r\bar{u}(n)u(p)$ in the limit of vanishing gluon mass μ^2. As a result, L^- remains finite in the limit as $\mu^2\to 0$ and, by the dimensional argument stated above, we expect L^- to be finite in the limit $q^2\to-\infty$. This reasoning can be confirmed by direct evaluation of Eq. (B12), which gives

$$\lim_{q^2\to-\infty}L^-(q^2,\omega)=(g_r{}^2/64\pi^2)\omega; \qquad \text{(B13)}$$

substituting into Eq. (B9) then gives

$$\lim_{q^2\to-\infty;\,m\to 0}f=g_r{}^2/16\pi^2, \qquad \text{(B14)}$$

in agreement with the $[\lambda^a,\lambda^b]$ term in Eq. (11).

To order $g_r{}^4$, we will not try to calculate the finite part of f, but only the part which diverges logarithmically as $q^2\to-\infty$. By our dimensional argument, this

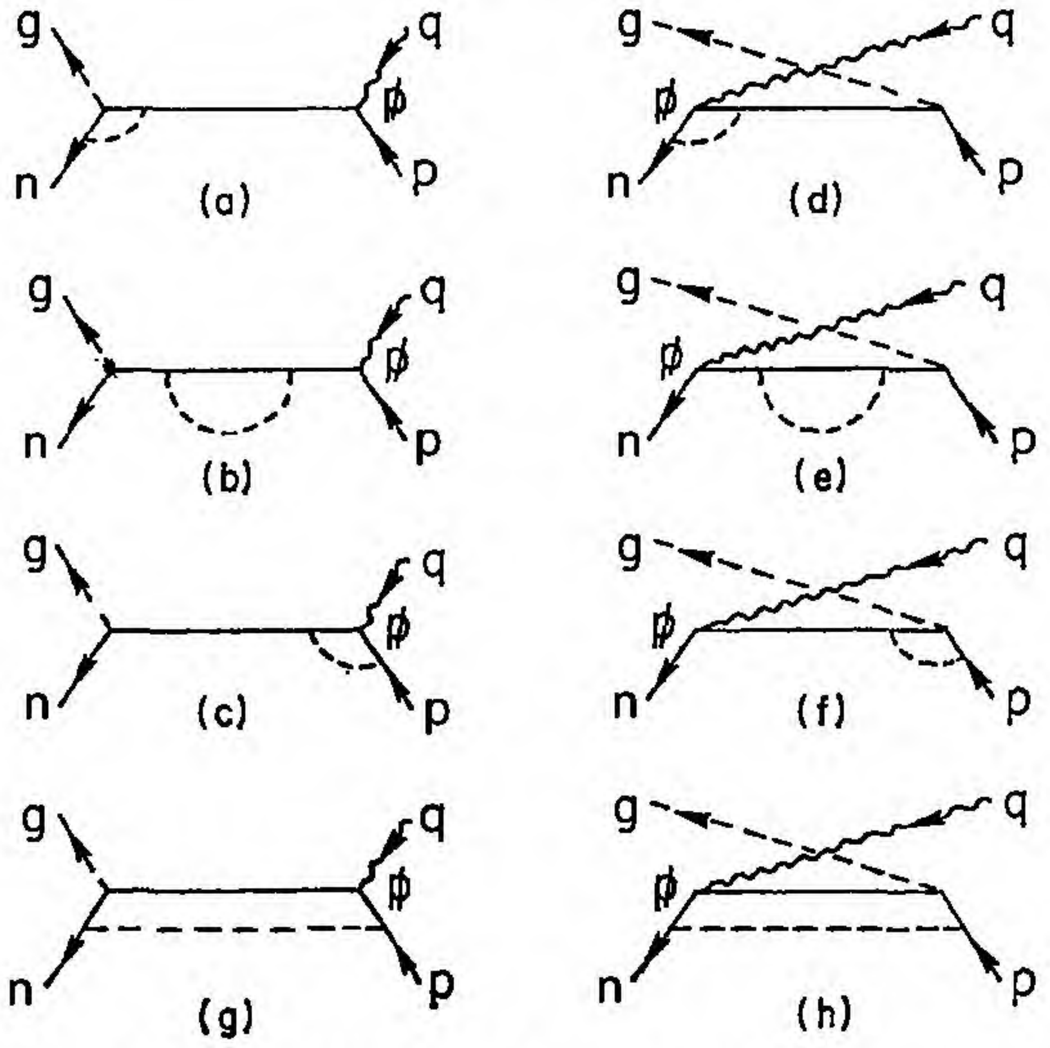


FIG. 7. Diagrams of order $g_r{}^3$ contributing to the "neutron" plus gluon intermediate state.

can be accomplished by isolating the part of f which diverges like $\ln\mu^2$ as $\mu^2\to 0$. There are four intermediate states which contribute in fourth order: (i) single "neutron", $N=n$ (Fig. 6); (ii) "neutron" plus one gluon, $N=n+g$ (Fig. 7); (iii) "neutron" plus two gluons, $N=n+g_1+g_2$ (Fig. 8); (iv) trident, $N=n+p+\bar{p}$, $n+n+\bar{n}$, or $p+p+\bar{n}$ (Fig. 9). The first three contribute only to L^-, while the trident intermediate state contributes to both L^+ and L^-. We consider the cases in turn.

(i) Single "neutron." The second-order part of $\langle n\,|\,p^{\tau}(J_r{}^1-iJ_r{}^2)\,|\,p\rangle$ is proportional to $\bar{u}(n)\tilde{\Lambda}(\gamma\cdot p;$

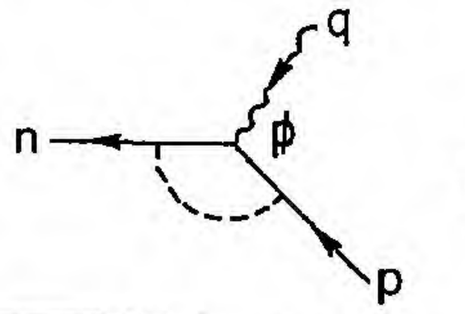


FIG. 6. Diagram of order $g_r{}^2$ contributing to the one "neutron" intermediate state.

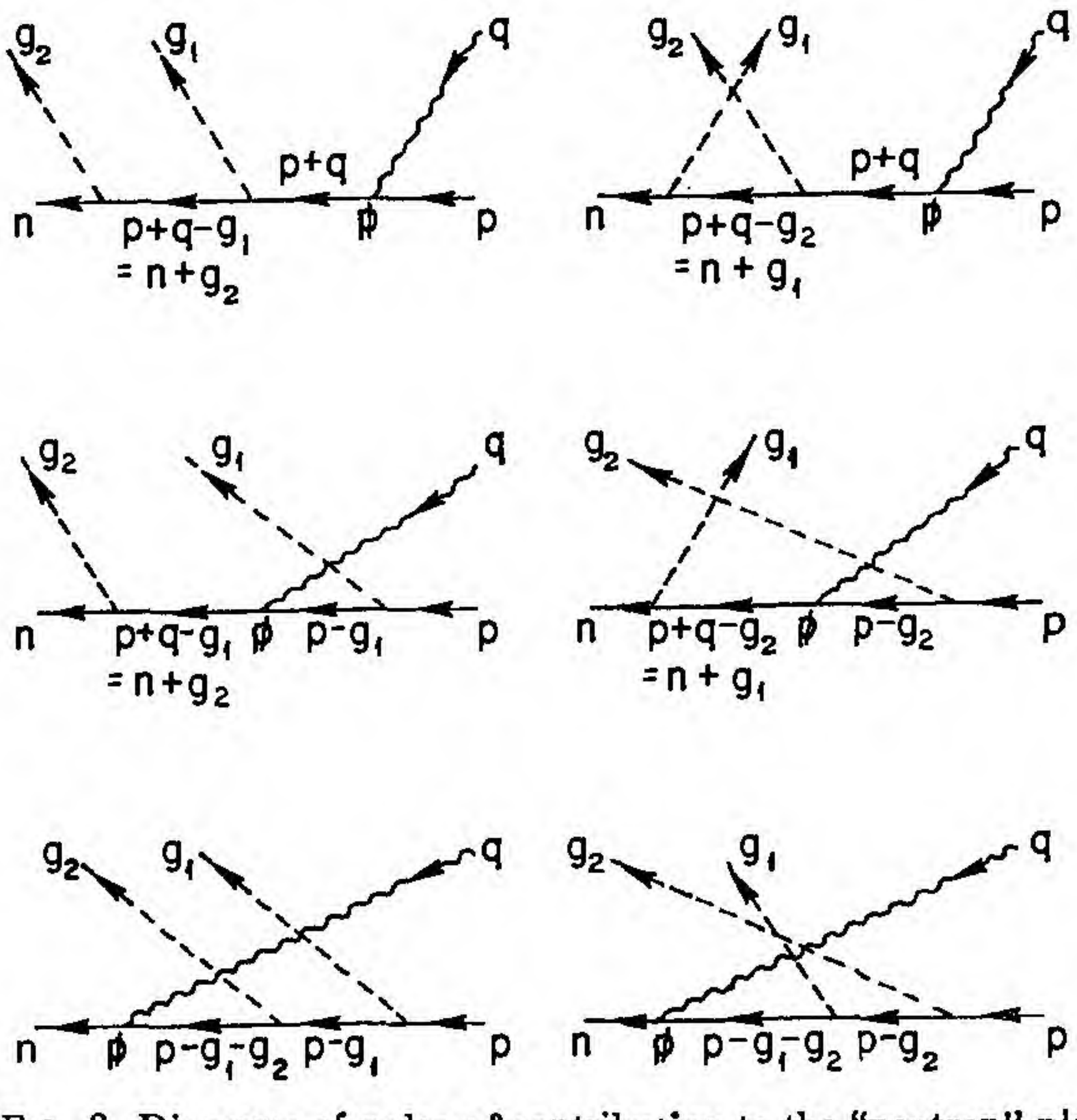


FIG. 8. Diagrams of order $g_r{}^2$ contributing to the "neutron" plus two gluon intermediate state.

[21] As we noted, the fermion mass m is zero. The factor m^2 in front of Eq. (B12) and subsequent equations just cancels a corresponding factor m^{-2} coming from our choice of spinor normalization.

TABLE IV. Phase-space regions and pieces of $|\mathfrak{M}|^2$ which actually make divergent contributions to Eq. (2.18). $n^s, g_1^s, \cdots$ denote the spatial components ($s=1,2,3$) of $n, g_1, \cdots$.

	Phase-space region	Piece of $\lvert\mathfrak{M}\rvert^2$
(4)	$g_1^s \parallel p^s$	$\lvert g_r^2\bar{u}(n)\{\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2)][1/(\gamma\cdot p-\gamma\cdot g_1)]\}u(p)\rvert^2$
(5)	$g_2^s \parallel p^s$	$\lvert g_r^2\bar{u}(n)\{\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2)][1/(\gamma\cdot p-\gamma\cdot g_2)]\}u(p)\rvert^2$
(7)	$g_1^s \parallel n^s$	$\lvert g_r^2\bar{u}(n)\{[1/(\gamma\cdot n+\gamma\cdot g_1)]\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_2)]\}u(p)\rvert^2$
(8)	$g_2^s \parallel n^s$	$\lvert g_r^2\bar{u}(n)\{[1/(\gamma\cdot n+\gamma\cdot g_2)]\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_1)]\}u(p)\rvert^2$

$n, p)u(p)$, where Λ is the renormalized vertex part. Using the fact that $p^2=0$, one sees from Eq. (A8) that $\tilde{\Lambda}(\gamma\cdot p; n, p)$ contains only a piece proportional to $\gamma\cdot p$ and a piece proportional to $(\gamma\cdot n)(\gamma\cdot p)(\gamma\cdot n)$, both of which vanish when sandwiched between the spinors. So the single- "neutron" contribution is zero.

(ii) "Neutron" plus one gluon. The "neutron"-plus- one- gluon contribution, in fourth order, arises from the interference of the first-order diagrams in Fig. 5 with the third-order diagrams in Fig. 7. The third-order diagrams are clearly of the same structure as the diagrams in Fig. 1, which we have already evaluated in our general treatment of the order $-g_r^2$ case. We note first that, because of the factor $\gamma\cdot p$, the contributions of Figs. 7(a) and 7(b) vanish. Thus, just as in the case of the first-order matrix element, the terms containing $(\gamma\cdot p+\gamma\cdot q)^{-1}\propto(2-\omega)^{-1}$ vanish, and

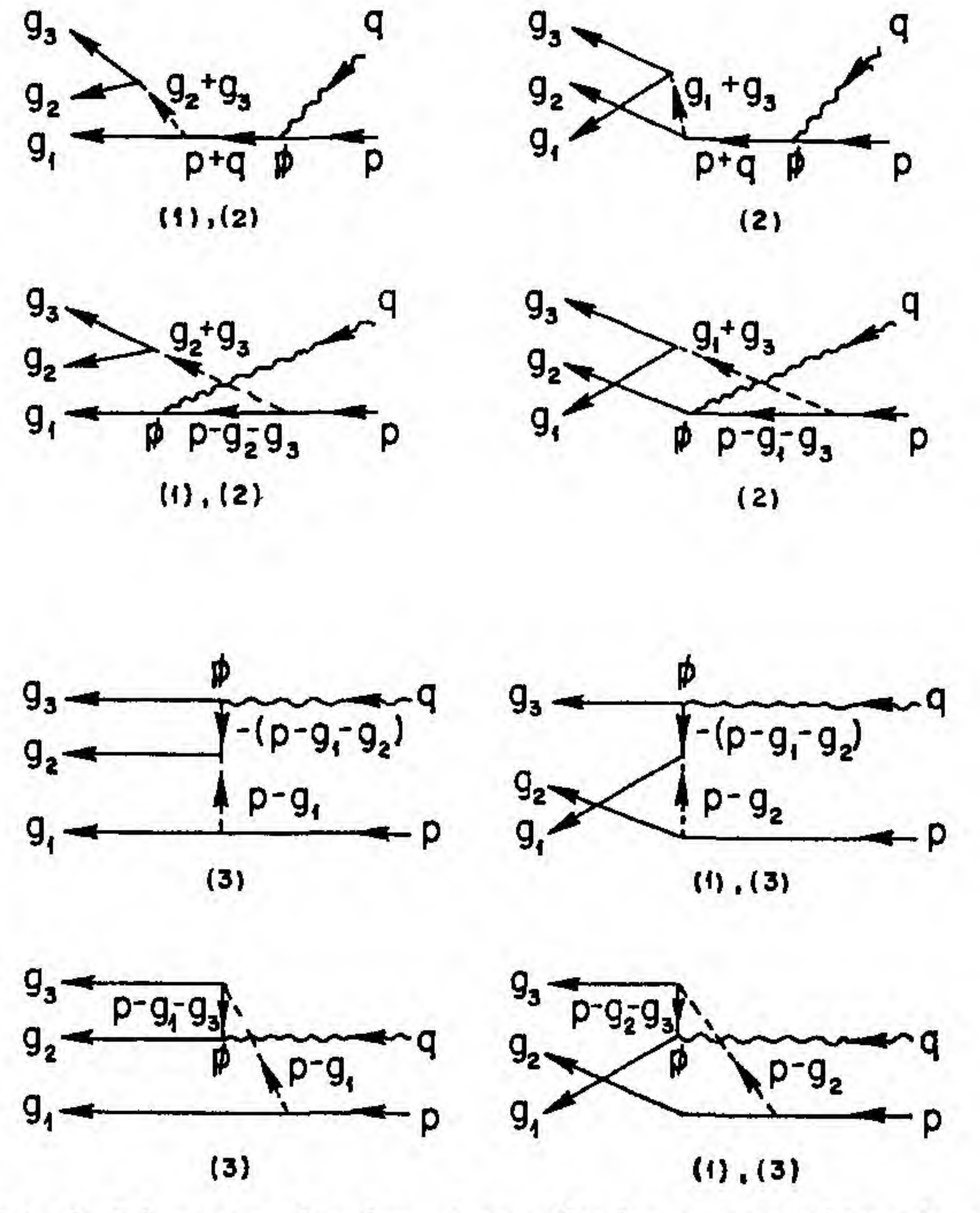


FIG. 9. Diagrams of order g_r^2 contributing to the trident intermediate states.

as a result the integral over ω' in Eq. (B9) converges, even for vanishing gluon mass μ^2. This means that any $\ln\mu^2$ singularities in f must result from $\ln\mu^2$ singularities in L^- itself.

To evaluate the contribution of Figs. 7(c)–7(f), we calculate the *renormalized* self-energy and vertex parts $\tilde{\Sigma}$ and $\tilde{\Lambda}$, by performing the usual mass and wave-function renormalizations on the unrenormalized quantities of Eqs. (A10). Note that the renormalized quantities contain no dependence on the cutoff λ, guaranteeing the validity of our dimensional arguments. In the treatment of the gluon vertex correction in Fig. 7(f), a subtlety arises. Instead of subtracting the vertex part at $g^2=\mu^2$, as required by the Watson-Lepore[22] convention, we subtract at $g^2=0$. The difference between the two methods of subtraction makes a contribution to L^- which is proportional to $\ln(m^2/\mu^2)$, but which, for fixed ω, is independent of q^2 and therefore can be dropped. This is the only place in the entire calculation where we encounter $\ln m^2$ terms and where the presence of a $\ln\mu^2$ term does not indicate the presence of a term $-\ln q^2$. When the gluon vertex part is subtracted at $g^2=0$, the $m\to 0$ limit is finite, and our usual dimensional argument applies. On substituting the expressions for $\tilde{\Sigma}$ and $\tilde{\Lambda}$ into the third-order matrix element, we find that the integration over the intermediate state $(n+g)$ variables is always convergent, so that $\ln\mu^2$ terms in L^- can *only arise* from the explicit $\ln\mu^2$ dependence of $\tilde{\Sigma}$ and $\tilde{\Lambda}$. We then find for the contributions of the various diagrams to L^-,

$$\begin{aligned}
&\lim_{\mu^2\to 0} L_{7(c)}^- = \text{finite},\\
&\lim_{\mu^2\to 0} L_{7(d)}^- = (g_r^2/4\pi)^2(\omega/64\pi^2)\ln\mu^2+\text{finite},\\
&\lim_{\mu^2\to 0} L_{7(e)}^- = -(g_r^2/4\pi)^2(\omega/64\pi^2)\ln\mu^2+\text{finite},\\
&\lim_{\mu^2\to 0} L_{7(f)}^- = -2(g_r^2/4\pi)^2(\omega/64\pi^2)\ln\mu^2+\text{finite}.
\end{aligned}\tag{B15}$$

Next, we must examine the contribution of the box diagrams of Figs. 7(g) and 7(h). We deal with these diagrams by writing them in Feynman parametrized form and substituting into the expression for L^-. For example, the contribution of Fig. 7(g) to L^- is pro-

[22] K. M. Watson and J. V. Lepore, Phys. Rev. 76, 1157 (1949).

portional to

$$\int_0^{\bar{\nu}_{\max}} d\bar{\nu} \int_0^1 x^2 dx \int_0^1 z\,dz \int_0^1 dy$$
$$\times \{(1/D_a^2)2x[(1-z+yz)\nu-yz\bar{\nu}][(1-x)\mu^2+2xz(1-y)\nu]+(2/D_a)[\nu[1-2x(1-z+yz)]+2xyz\bar{\nu}]\}, \quad \text{(B16)}$$

$$D_a=\mu^2[x-1+x^2yz(1-z)]+x(1-x)(1-z)(p+q)^2-x^2(1-y)z[2(1-z+yz)\nu-2yz\bar{\nu}-(1-z)(p+q)^2],$$
$$\nu=p\cdot q, \quad \bar{\nu}=p\cdot g.$$

For general values of $q\cdot p$ and q^2, a singularity of Eq. (B16) at $\mu^2=0$ can only arise from the integration end points $\bar{\nu}=0$, $x=0$, $x=1$, $\cdots$, $y=1$.[23] A careful analysis of the behavior of Eq. (B16) at these end points in all possible combinations shows that there is no $\ln\mu^2$ term as $\mu^2\to 0$. A similar analysis yields the same result for Fig. 7(h), so we get, finally,

$$\lim_{\mu^2\to 0} L_{7(g)}^- = \lim_{\mu^2\to 0} L_{7(h)}^- = \text{finite}. \quad \text{(B17)}$$

This completes our analysis of the "neutron" plus one gluon intermediate state.

(iii) "Neutron" plus two gluons. The "neutron" plus two gluon contribution arises from the square of the second-order matrix element corresponding to the diagrams in Fig. 8. We have

$$L_{2\ \text{gluon}}^- = \frac{-m\omega^2}{q^2}(2\pi)^3\tfrac{1}{4}\sum_{\text{spin}(p)}\sum_{\text{spin}(n)}\int\frac{d^3n}{(2\pi)^3}\frac{m}{n^0}\frac{1}{2}\int\frac{d^3g_1}{(2\pi)^3}\frac{1}{2g_1^0}\int\frac{d^3g_2}{(2\pi)^3}\frac{1}{2g_2^0}\delta^4(p+q-n-g_1-g_2)\,|\mathfrak{M}|^2, \quad \text{(B18)}$$

with

$$\mathfrak{M}=g_r^2\bar{u}(n)\left(\frac{1}{\gamma\cdot n+\gamma\cdot g_2}\gamma\cdot p\frac{1}{\gamma\cdot p-\gamma\cdot g_1}+\frac{1}{\gamma\cdot n+\gamma\cdot g_1}\gamma\cdot p\frac{1}{\gamma\cdot p-\gamma\cdot g_2}+\gamma\cdot p\frac{1}{\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2}\frac{1}{\gamma\cdot p-\gamma\cdot g_1}+\gamma\cdot p\frac{1}{\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2}\frac{1}{\gamma\cdot p-\gamma\cdot g_2}\right)u(p). \quad \text{(B19)}$$

Only four terms appear in $\mathfrak{M}$ because the contributions of the two diagrams on the first line of Fig. 8 are proportional to $(\gamma\cdot p+\gamma\cdot q)^{-1}\gamma\cdot pu(p)$, and therefore vanish. Just as before, this means that the integral over ω' in Eq. (B9) converges, and any $\ln\mu^2$ behavior in $f_{2\ \text{gluon}}$ must originate in $L_{2\ \text{gluon}}^-$ itself. Possible divergences in $L_{2\ \text{gluon}}^-$ as $\mu^2\to 0$ arise from the eight regions of three-particle phase space listed in Table III, where denominators in the matrix element of Eq. (B19) vanish. To extract the divergent part, we make a careful study of the behavior of the integral of Eq. (B18) in each of the eight regions. In this connection, the following simple inequality is very useful: Let p be a null vector and let $Q(=g_1, g_2, g_1+g_2)$ be timelike with $p^0>0$ and $Q^0>0$. Then we may write

$$(\gamma\cdot p)(\gamma\cdot Q)=p\cdot Q+\tfrac{1}{2}\gamma_\alpha\gamma_\beta T^{\alpha\beta},$$
$$T^{\alpha\beta}=p^\alpha Q^\beta-p^\beta Q^\alpha, \quad \text{(B20)}$$

with the following simple bounds on $T^{\alpha\beta}$:

$$|T^{AB}|\le[4p^0Q^0p\cdot Q]^{1/2}, \quad A,B,=1,2,3$$
$$|T^{A0}|\le[2(p\cdot Q)^2+4p^0Q^0p\cdot Q]^{1/2}. \quad \text{(B21)}$$

In other words, for small $p\cdot Q$, the γ-matrix product $(\gamma\cdot p)(\gamma\cdot Q)$ is always bounded by $(p\cdot Q)^{1/2}$. Application of this inequality shows that many of the potentially divergent phase-space regions actually make a finite contribution to Eq. (B18), and that the only divergent contributions come from the phase-space regions and pieces of $|\mathfrak{M}|^2$ shown in Table IV. Evaluation of the spin sums and phase-space integrals show that regions (4) and (5) each make a contribution to L^- of

$$-\tfrac{1}{2}(g_r^2/4\pi)^2(\omega/64\pi^2)[\ln(\tfrac{1}{2}\omega)+(2/\omega)-1]\ln\mu^2+\text{finite}, \quad \text{(B22)}$$

while regions (7) and (8) each make a contribution of

$$-\tfrac{1}{4}(g_r^2/4\pi)^2(\omega/64\pi^2)\ln\mu^2, \quad \text{(B23)}$$

giving a total of

$$\lim_{\mu^2\to 0} L_{2\ \text{gluon}}^- = -(g_r^2/4\pi)^2(\omega/64\pi^2)[\ln(\tfrac{1}{2}\omega)+(2/\omega)-\tfrac{1}{2}]\times\ln\mu^2+\text{finite}. \quad \text{(B24)}$$

(iv) Trident. The three trident contributions arise from the squares of the second-order matrix elements corresponding to the diagrams of Fig. 9. In Table V we list the momentum labeling for each of the three states and indicate to which L it contributes. The

TABLE V. Four-momentum labeling for trident production.

Trident state	Four-momentum label g_1	g_2	g_3	Contributes to
(1)	n	p	$\bar{p}$	L^-
(2)	n	n	$\bar{n}$	L^-
(3)	p	p	$\bar{n}$	L^+

[23] T. Kinoshita, J. Math. Phys. 3, 650 (1952).

TABLE VI. Phase-space regions and pieces of $|\mathfrak{M}^{(1,2)}|^2$ which actually make divergent contributions to Eq. (3.25). g_1^s, g_2^s, and g_3^s denote the spatial components ($s=1,2,3$) of g_1, g_2, and g_3.

Phase-space region	Piece of $\lvert\mathfrak{M}\rvert^2$	Occurs in
$g_2^s \parallel g_3^s$	$\lvert g_r^2\bar{u}(g_1)\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_2-\gamma\cdot g_3)]u(p)[1/((g_2+g_3)^2-\mu^2)]\bar{u}(g_2)v(g_3)\rvert^2$	$\lvert\mathfrak{M}^{(1)}\rvert^2$, $\lvert\mathfrak{M}^{(2)}\rvert^2$
$g_1^s \parallel g_3^s$	$\lvert g_r^2\bar{u}(g_2)\gamma\cdot p[1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_3)]u(p)[1/((g_1+g_3)^2-\mu^2)]\bar{u}(g_1)v(g_3)\rvert^2$	$\lvert\mathfrak{M}^{(2)}\rvert^2$

matrix element for state (j) ($j=1, 2, 3$) receives contributions from only those diagrams in Fig. 9 which are labeled below with (j). We find [the factors of $\frac{1}{2}$ in Eq. (B26) are statistical]

$$L^{\mp}{}_{\text{trident}}=\frac{-m\omega^2}{q^2}(2\pi)^3\tfrac{1}{4}\sum_{\text{spin}(p)}\;\sum_{\text{spin}(g_1,g_2,g_3)}\int\frac{d^3g_1}{(2\pi)^3}\frac{m}{g_1^0}\int\frac{d^3g_2}{(2\pi)^3}\frac{m}{g_2^0}\int\frac{d^3g_3}{(2\pi)^3}\frac{m}{g_3^0}\delta^4(p+q-g_1-g_2-g_3)\,|\mathfrak{M}^{\mp}|^2, \quad \text{(B25)}$$

with

$$|\mathfrak{M}^-|^2=|\mathfrak{M}^{(1)}|^2+\tfrac{1}{2}|\mathfrak{M}^{(2)}|^2,\qquad |\mathfrak{M}^+|^2=\tfrac{1}{2}|\mathfrak{M}^{(3)}|^2, \quad \text{(B26)}$$

$$\begin{aligned}\mathfrak{M}^{(1)}=g_r^2\{&\bar{u}(g_1)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_2-\gamma\cdot g_3)^{-1}u(p)[(g_2+g_3)^2-\mu^2]^{-1}\bar{u}(g_2)v(g_3)\\&+\bar{u}(g_2)u(p)[(p-g_2)^2-\mu^2]^{-1}\bar{u}(g_1)[-1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2)]\gamma\cdot pv(g_3)\\&+\bar{u}(g_2)u(p)[(p-g_2)^2-\mu^2]^{-1}\bar{u}(g_1)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_2-\gamma\cdot g_3)^{-1}v(g_3)\},\\ \mathfrak{M}^{(2)}=g_r^2\{&\bar{u}(g_1)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_2-\gamma\cdot g_3)^{-1}u(p)[(g_2+g_3)^2-\mu^2]^{-1}\bar{u}(g_2)v(g_3)\\&+\bar{u}(g_2)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_3)^{-1}u(p)[(g_1+g_3)^2-\mu^2]^{-1}\bar{u}(g_1)v(g_3)\}, \quad \text{(B27)}\\ \mathfrak{M}^{(3)}=g_r^2\{&\bar{u}(g_1)u(p)[(p-g_1)^2-\mu^2]^{-1}\bar{u}(g_2)[-1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2)]\gamma\cdot pv(g_3)\\&+\bar{u}(g_2)u(p)[(p-g_2)^2-\mu^2]^{-1}\bar{u}(g_1)[-1/(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_2)]\gamma\cdot pv(g_3)\\&+\bar{u}(g_1)u(p)[(p-g_1)^2-\mu^2]^{-1}\bar{u}(g_2)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_1-\gamma\cdot g_3)^{-1}v(g_3)\\&+\bar{u}(g_2)u(p)[(p-g_2)^2-\mu^2]^{-1}\bar{u}(g_1)\gamma\cdot p(\gamma\cdot p-\gamma\cdot g_2-\gamma\cdot g_3)^{-1}v(g_3)\}.\end{aligned}$$

The two diagrams on the first line of Fig. 9 make no contribution to the matrix elements since they contain the factor $(\gamma\cdot p+\gamma\cdot q)^{-1}\gamma\cdot pu(p)=0$, and as before, this means that divergences in f_{trident} must originate in $L^{\mp}{}_{\text{trident}}$ themselves. Potential divergences in $L^{\mp}{}_{\text{trident}}$ are associated with special regions of three-body phase space where denominators in Eq. (B27) vanish. In studying the actual behavior of Eq. (B25) in these regions, we use the inequality of Eq. (B21) and the estimates

$$\begin{aligned}|\bar{u}(g_{1,2})u(p)|&\propto(g_{1,2}\cdot p)^{1/2}\quad\text{as } g_{1,2}\cdot p\to 0,\\ |\bar{u}(g_{1,2})v(g_3)|&\propto(g_{1,2}\cdot g_3)^{1/2}\quad\text{as } g_{1,2}\cdot g_3\to 0. \quad \text{(B28)}\end{aligned}$$

We find that most of the dangerous phase-space regions actually give finite results in the $\mu^2\to 0$ limit, with logarithmic divergences coming from the regions of phase space and pieces of $|\mathfrak{M}^{(1,2)}|^2$ shown in Table VI. Evaluation of the spin sums and phase-space integrals gives the result

$$\lim_{\mu^2\to 0}L^+{}_{\text{trident}}=\text{finite},$$

$$\lim_{\mu^2\to 0}L^-{}_{\text{trident}}=-4(g_r^2/4\pi)^2(\omega/64\pi^2)\ln\mu^2+\text{finite}, \quad \text{(B29)}$$

with $\frac{3}{4}$ of this result coming from the phase-space region $g_2^s \parallel g_3^s$ and $\frac{1}{4}$ from the region $g_1^s \parallel g_3^s$.

This completes our analysis of intermediate states which contribute in order g_r^4. Adding up the contributions from Eqs. (B15), (B24), and (B29), we find, for the total fourth-order contribution,

$$\lim_{\mu^2\to 0}L^+(q^2,\omega)=\text{finite},$$

$$\lim_{\mu^2\to 0}L^-(q^2,\omega)=-(g_r^2/4\pi)^2(\omega/64\pi^2) \quad \text{(B30)}$$
$$\times[\ln(\tfrac{1}{2}\omega)+(2/\omega)+\tfrac{11}{2}]\ln\mu^2+\text{finite},$$

which, by our dimensional argument, implies that

$$\lim_{q^2\to-\infty}L^+(q^2,\omega)=\text{finite},$$

$$\lim_{q^2\to-\infty}L^-(q^2,\omega)=(g_r^2/4\pi)^2(\omega/64\pi^2) \quad \text{(B31)}$$
$$\times[\ln(\tfrac{1}{2}\omega)+(2/\omega)+\tfrac{11}{2}]\ln(q^2/\mu^2)+\text{finite}.$$

Substituting this result into Eqs. (B6) and (B9) yields the fourth-order Bjorken limit quoted in Eq. (12) of the text.[24]

[24] A fourth-order calculation of the longitudinal cross section in the inequivalent limit in which $|q^2|$ and ω^{-1} simultaneously become large has been given recently by H. Cheng and T. T. Wu, Phys. Rev. Letters 22, 1409 (1969).

Excerpt from S. L. Adler, Anomalies in Ward Identities and Current Commutation Relations, in *Local Currents and Their Applications, Proceedings of an Informal Conference*, D. H. Sharp and A. S. Wightman, eds. (North-Holland, Amsterdam and American Elsevier, New York, 1974). Reprinted with permission from Elsevier.

2.4. *Questions raised by the breakdown of the BJL limit*

Experimentally, Bjorken scaling works very well and $\sigma_S/\sigma_T = 0.18 \pm 0.05$, i.e. the longitudinal cross section *is* small. So renormalized perturbation theory seems to be a bad guide here. This state of affairs raises several questions.

(i) Is it only the perturbation expansion that is at fault, or *does the trouble lie in local field theory itself?* Bitar and Khuri [4] have studied the BJL limit using only analyticity and positivity. They find that class I intermediate state (fig. 10) violate the BJL limit for space-space commutators, but cannot rule out a cancelling contribution from class II intermediate states (fig. 11).

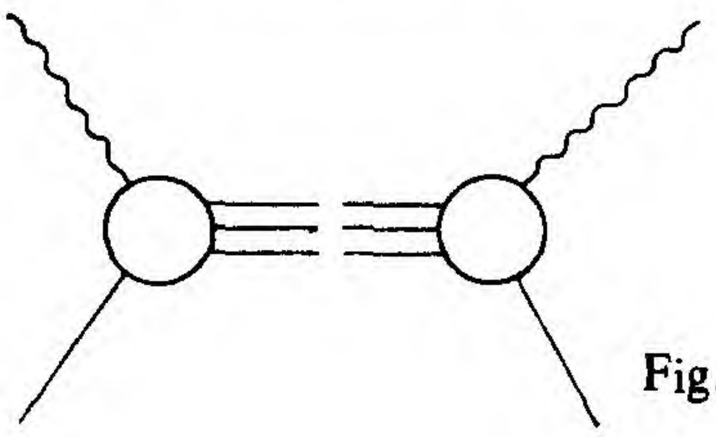
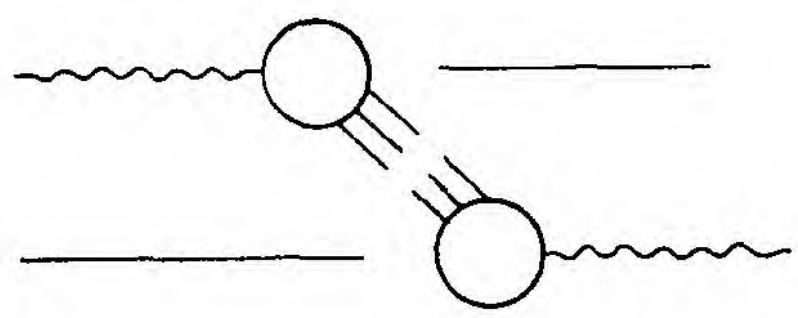


Fig. 10



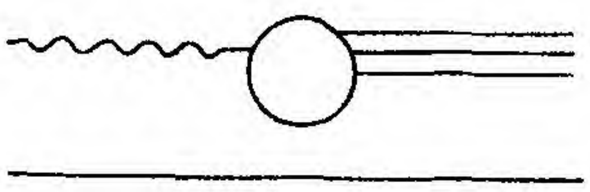
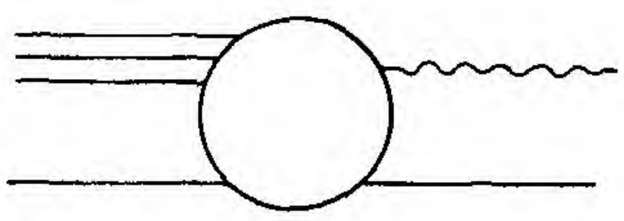


Fig. 11



(ii) Can one make a consistent calculational scheme in which Bjorken limits, the Callan-Gross relation and scaling are all valid? This is a *real challenge* to theorists. The Lee-Wick theory, for example, doesn't do the job: The BJL limits are satisfied but complex singularities change dispersion relations in such a way that the Callan-Gross relation is still violated. Perhaps a successful approach would involve summation of perturbation theory graphs plus use of the Gell-Mann-Low eigenvalue condition (see sect. 3).

(iii) The same questions apply to light-cone algebra, which is basically the BJL limit in the $p_0 \to \infty$ frame, as in the Callan-Gross derivation of their sum rules.

3. Anomalous scaling

Consider a field theory with a dimensionless coupling constant. When all energies become much greater than particle masses, one naively expects the n-point functions to scale – i.e. to become *mass-independent* apart from an overall factor. In this section we discuss the formal theory of scaling[5] and its breakdown in field theory.

3.1. *Formal theory of scaling*

The infinitesimal generator of dilations, δ_D, transforms coordinates as follows:

$$\delta_D x^\mu = -x^\mu \tag{49}$$

Under the transformations (49), a field φ transforms as

$$\delta_D \varphi = (x_\lambda \partial^\lambda + d)\varphi , \tag{50}$$

Where d is the "scale dimension" of φ. Scale invariance for renormalizable field theories results if we take

$$\begin{aligned} d &= 1 \quad \text{for bosons,} \\ d &= \frac{3}{2} \quad \text{for fermions.} \end{aligned} \tag{51}$$

In simple canonical field theories it is possible to find a conserved, symmetric energy-momentum tensor $\theta_{\mu\nu}$ which can be used to define a "dilation current" D_μ:

$$D_\mu = x^\nu \theta_{\mu\nu} \,. \tag{52}$$

The energy-momentum tensor $\theta_{\mu\nu}$ is constructed so that its trace is proportional to the mass terms in the Lagrangian. Thus .

$$\partial^\mu D_\mu = \theta^\mu_\mu = \sum_j \frac{\partial \mathcal{L}_m}{\partial \varphi_j} d_j \varphi_j - 4\mathcal{L}_m \,, \tag{53}$$

where $\mathcal{L}_m$ = mass term in the Lagrangian (the only term which breaks scale invariance). Therefore, the dilation current D_μ is conserved when the theory is scale-invariant. Even when D_μ is not conserved, the "charge"

$$D(t) = \int d^3 x \, D_0(x, t) \tag{54}$$

acts as a generator of dilations:

$$i[D(t), \varphi_j(x, t)] = (x_\lambda \partial^\lambda + d_j)\varphi_j(x, t). \tag{55}$$

The relationship

$$\partial^\mu D_\mu = \theta_\mu{}^\mu \tag{56}$$

is the scale-invariance analog of PCAC. Like PCAC, it can be used to derive low-energy theorems; in the present case for the emission of gravitons in an arbitrary process.

Now consider a single scalar-meson field φ with a φ^4 self-interaction. Let $G(p)$ be the renormalized propagator and $\Gamma(p, q)$ the θ^μ_μ vertex function. From the usual definitions of these quantities we have

$$G(p)\,\Gamma(p, q) G(p + q)\big|_{q=0}$$

$$= \int d^4 x\, d^4 y\, e^{iq\cdot x} e^{ip\cdot y} \langle 0 | T^*(\varphi(y)\varphi(0)\theta^\mu_\mu(x)) | 0 \rangle \big|_{q=0} \tag{57}$$

$$= \int d^4 x\, d^4 y\, e^{iq\cdot x} e^{ip\cdot y} \{ \partial^\mu T^* (\varphi(y)\varphi(0)D_\mu) - \delta(x_0 - y_0)[D_0(x), \varphi(y)]\varphi(0)$$

$$- \varphi(y)\delta(x_0)[D_0(x), \varphi(0)] \} \big|_{q=0} \,. \tag{58}$$

Integration by parts shows that the first term on the r.h.s. is zero when $q = 0$, so eq. (58) becomes

$$G(p)\Gamma(p,0)G(p) = -\int d^4y e^{ip\cdot y}\{\delta(x_0-y_0)\langle 0|[D(x_0),\varphi(y)]\varphi(0)|0\rangle$$
$$+\ \delta(x_0)\langle 0|\varphi(y)[D(x_0),\varphi(0)]|0\rangle\}\,. \tag{59}$$

The quantity in curly brackets on the r.h.s. can be related back to $G(p)$ using the dilation generator commutator [eq. (55)], and in this way one gets

$$-i\Gamma(p,0) = p^{\nu}\partial_{\nu}G^{-1}(p) + (2d-4)G^{-1}(p). \tag{60}$$

Recall that $\Gamma(p, 0)$ is the three-point function shown in fig. 12, Weinberg's theorem says that $\Gamma(p, 0)\sim$ polynomial in $\log p^2$ as $p^2 \to -\infty$. Thus, neglecting

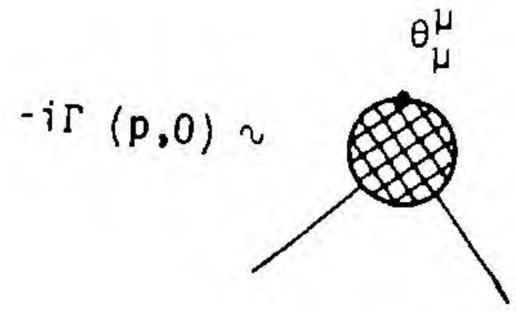


Fig. 12

$\Gamma(p, 0)$ as $p \to \infty$ in eq. (60), we find

$$p^{\nu}\partial_{\nu}G_{\infty}^{-1}(p) = (4-2d)G_{\infty}^{-1}(p)\,. \tag{61}$$

This implies that G_{∞}^{-1} satisfies the scaling law

$$\begin{aligned} G_{\infty}^{-1} &= A(p^2)^{2-d} \\ &= Ap^2 \qquad \text{when } d = 1. \end{aligned} \tag{62}$$

One sees from this that the neglect of $\Gamma(p, 0)$ in eq. (60) was justified.

Unfortunately, we know that the scaling behavior for G_{∞}^{-1} predicted by the above chain of formal arguments is not correct in renormalized perturbation theory, where we find that

$$G_{\infty}^{-1} = p^2 \times [\text{power series in} \log p^2]\,. \tag{63}$$

The reason for the breakdown of the formal theory of scaling is not hard to find.

To make the formal manipulations valid, we need to put in regulators. But the regulators involve *large masses*, which necessarily break scale invariance. The trouble arises because $(\theta_\mu^\mu)^{\mathrm{regl}}$ does *not* go to zero as the regulator masses go to infinity. This behavior is of course analogous to what one finds when one looks at the VVA Ward identity from the regulator viewpoint.

3.2. *Correct scaling relations: The Callan-Symanzik equations*

Straightforward dimensional arguments require that $G^{-1}(p)$ depend on the particle mass as follows

$$G^{-1}(p) = \mu^2 G^{-1}[p^2/\mu^2] , \tag{64}$$

so that we have the identity

$$p_\nu \frac{\partial}{\partial p_\nu} G^{-1}(p) = 2G^{-1}(p) - \mu \frac{\mathrm{d}}{\mathrm{d}\mu} G^{-1}(p). \tag{65}$$

The naive scaling relation, eq. (60), can thus be written

$$-i\Gamma(p, 0) = (-\mu \frac{\partial}{\partial\mu} + 2\gamma) G^{-1}(p), \qquad G^{-1}(p), \tag{66}$$

where $\gamma = d - 1$ (= 0 when the "scale dimension" d equals the naive canonical dimension = 1).

Callan and Symanzik [6] have shown that a correct version of the scaling law (66) is given by

$$-i\Gamma(p,0) = (-\mu \frac{\partial}{\partial\mu} + 2\gamma(\lambda) + \beta(\lambda) \frac{\partial}{\partial\lambda}) G^{-1}(p) \tag{67}$$

where λ is the renormalized coulping constant, $\gamma(\lambda) = d(\lambda) - 1$ and $d(\lambda)$ is the "anomalous" scaling dimension of the theory. *If* the β term were zero, one would find

$$G_\infty^{-1} = A(p^2)^{2-d(\lambda)} = A(p^2)^{1-\gamma(\lambda)} , \tag{68}$$

i.e., scaling with anomalous dimension $d(\lambda)$.

A simple scaling law like eq. (68) does not result from eq. (67) with $\beta \neq 0$. The $p \to \infty$ limit of eq. (67) then becomes

$$0 \sim \left(-\mu \frac{\partial}{\partial \mu} + 2\gamma(\lambda) + \beta(\lambda) \frac{\partial}{\partial \lambda}\right) G_{\infty}^{-1}(p), \tag{69}$$

which can be integrated and gives only the much less restrictive asymptotic predictions of renormalization group theory.

3.3. *Remarks*

We conclude with several comments on the foregoing results.

(i) The statement $\beta(\lambda) = 0$ is closely related to the Gell-Mann-Low eigenvalue condition*.

(ii) If $\beta(\lambda) = 0$, with a simple zero λ_0, then

$$G_{\infty}^{-1}(p) = A(p^2)^{1-\gamma(\lambda_0)} \tag{70}$$

In other words, the asymptotic behavior is determined by the bare coupling constant λ_0, independent of the value of the renormalized coupling constant λ[7].

(iii) Several models with $\beta = 0$ show scaling properties. For example, this is the case for both the Johnson-Baker-Willey model[8] of quantum electrodynamics and for the Thirring model[9] (with massless or massive fermions). Note that neither of these models has a vacuum-polarization structure, which for a fermion theory implies $\beta \equiv 0$.

(iv) The Callan-Symanzik equation (67) can be used to prove the momentum space version of Wilson's operator product expansion (in perturbation theory), and this can be used to study the anomalous BJL limit[10].

(v) An interesting question is whether the Callan-Symanzik relations can be used to do graph summations for objects more complicated than propagators, for example, for fermion inelastic structure functions[11].

References

[1] S.L. Adler, Perturbation theory anomalies, in Lectures on elementary particles and quantum field theory, ed. S. Deser, M. Grisaru and H. Pendleton (M.I.T. Press, Cambridge, Mass., 1970).

[2] A Zee, Phys. Rev. Letters 29 (1972) 1198; B. Schroer and J. Lowenstein, Phys. Rev. D7 (1972) 1929.

[3] J. Wess and B. Zumino, Phys. Letters 37B (1971) 95.

* For a discussion see ref. [7] and references cited therein.

[4] K. Bitar and N.N. Khuri, Phys. Rev.D3 (1971) 462.
[5] S. Coleman and R. Jackiw, Ann. of Phys. 67 (1971) 522.
[6] C.G. Callan, Phys. Rev. D2 (1970) 1541; K. Symanzik, Comm. Math. Phys. 18 (1970) 227.
[7] S.L. Adler, Phys. Rev. D5 (1972) 3021.
[8] S.L. Adler and W.A. Bardeen, Phys. Rev. D4 (1971) 3045.
[9] A.H. Mueller and T.L. Trueman, Phys. Rev. D4 (1971) 1635.
[10] C.G. Callan, Phys. Rev. 12 (1972) 3202.
[11] N. Christ, B. Hasslacher and A.H. Mueller, Phys. Rev. D6 (1972) 3543.

In reference [10], Phys. Rev. 12 should read Phys. Rev. D5.

PHYSICAL REVIEW D VOLUME 15, NUMBER 6 15 MARCH 1977

Energy-momentum-tensor trace anomaly in spin-1/2 quantum electrodynamics

Stephen L. Adler*
The Institute for Advanced Study, Princeton, New Jersey 08540
and Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87544

John C. Collins†
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540

Anthony Duncan*
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 10 June 1976; revised manuscript received 1 November 1976)

We relate the energy-momentum-tensor trace anomaly in spin-1/2 quantum electrodynamics to the functions $\beta(\alpha)$, $\delta(\alpha)$ defined through the Callan-Symanzik equations, and prove finiteness of $\theta_{\mu\nu}$ when the anomaly is taken into account.

I. INTRODUCTION

Spin-$\frac{1}{2}$ quantum electrodynamics, characterized by the Lagrangian density[1]

$$\mathcal{L}_{\rm inv}(x)=\bar{\psi}(x)(i\gamma\cdot\partial-m_0)\psi(x)-\tfrac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x)-e_0\bar{\psi}(x)\gamma_\mu\psi(x)A^\mu(x)\,, \tag{1.1}$$

is one of the simplest field theory models in which to study anomalies. The axial-vector divergence anomaly in this theory has been extensively analyzed[2]; we wish in this note to discuss some properties of the energy-momentum-tensor trace anomaly.[3] Taking for the energy-momentum tensor $\theta_{\mu\nu}$ the symmetric form

$$\begin{aligned}\theta_{\mu\nu}&=\theta^{e\gamma}_{\mu\nu}+\theta^{\gamma}_{\mu\nu}\,,\\ \theta^{\gamma}_{\mu\nu}&=\tfrac{1}{4}\eta_{\mu\nu}F_{\lambda\sigma}F^{\lambda\sigma}-F_{\lambda\mu}F^{\lambda}{}_{\nu}\,,\\ \theta^{e\gamma}_{\mu\nu}&=\tfrac{1}{4}i[\bar{\psi}\gamma_\mu(\vec{\partial}_\nu+ie_0A_\nu)\psi+\bar{\psi}\gamma_\nu(\vec{\partial}_\mu+ie_0A_\mu)\psi\\ &\quad-\bar{\psi}(\overleftarrow{\partial}_\nu-ie_0A_\nu)\gamma_\mu\psi-\bar{\psi}(\overleftarrow{\partial}_\mu-ie_0A_\mu)\gamma_\nu\psi]\,,\end{aligned} \tag{1.2}$$

a simple application of the equations of motion gives the so-called "naive" trace formula

$$\theta_\mu{}^\mu=m_0\bar{\psi}\psi\,. \tag{1.3}$$

As has been shown by the authors of Ref. 3, Eq. (1.3) is not correct as it stands, but instead must be modified by the addition of an anomalous term[4] proportional to $Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}$. Our aim in this paper is to derive an explicit formula for the trace anomaly, valid to all orders in perturbation theory, expressed in terms of the functions $\beta(\alpha)$ and $\delta(\alpha)$ of the fine-structure constant defined through the Callan-Symanzik equations.

In Sec. II we give a simple heuristic derivation of our result, which, as we shall see, is most naturally written in terms of a subtracted operator $N[F_{\lambda\sigma}F^{\lambda\sigma}]$. There, we will be thinking in terms of using massive regulator fields. Some related details are given in the appendixes.

Then in Sec. III we will give a more careful derivation using normal-product methods[5] and dimensional regularization.[6] In n space-time dimensions, we have

$$\theta_\mu{}^\mu=-(n-4)\mathcal{L}_{\rm inv}-3(\tfrac{1}{2}i\bar{\psi}\overleftrightarrow{\not{\partial}}\psi-m_0\bar{\psi}\psi)+m_0\bar{\psi}\psi\,. \tag{1.4}$$

The anomaly is the term $-(n-4)\mathcal{L}_{\rm inv}$, which would vanish if $\mathcal{L}_{\rm inv}$ were finite. We wish to express the anomaly in terms of renormalized operators.

Our derivation will give as a byproduct a proof that $\theta_{\mu\nu}$ as defined by Eq. (1.2) is finite to all orders of perturbation theory even when the trace anomaly is taken into account. The earlier proof by Callan, Coleman, and Jackiw[7] is incomplete, while the one by Freedman, Muzinich, and Weinberg[8] is not directly applicable to our case.

II. HEURISTIC DERIVATION

The heuristic derivation is obtained by writing down an operator formula for the trace equation and then determining the unknown coefficients appearing in this formula by studying its electron-to-electron and vacuum-to-two-photon matrix elements. As our initial operator ansatz let us write the most general linear combination of gauge-invariant scalar C-even operators with the correct dimensionality,

$$\begin{aligned}\theta_\mu{}^\mu=C_1m_0\bar{\psi}\psi+C_2Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}&\\ +C_3\tfrac{1}{2}i[\bar{\psi}\gamma\cdot(\vec{\partial}+ie_0A)\psi-\bar{\psi}\gamma\cdot(\overleftarrow{\partial}-ie_0A)\psi&\\ -2m_0\bar{\psi}\psi]\,.&\end{aligned} \tag{2.1}$$

The coefficient of C_3 is formally zero by use of the equations of motion; it represents a discontinuous contribution which is present at zero momentum transfer, but which vanishes for nonzero mo-

mentum transfers, and hence does not contribute to physical matrix elements. The precise structure of this term will be determined in Sec. III, but we will ignore it in the heuristic discussion which follows. Focussing on the first two terms, it is easy to see that either C_1 or C_2 is infinite, or Eq. (2.1) cannot be correct as it stands. The reason is that both $\theta_\mu{}^\mu$ and $m_0\bar{\psi}\psi$ are finite operators[9] (that is, their matrix elements are made finite by the usual electron and photon wave-function renormalizations), whereas a simple calculation shows that the *lowest-order* diagrams (illustrated in Fig. 1) contributing to the electron-to-electron and vacuum-to-two-photon matrix elements of $Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}$ are logarithmically divergent, and hence cannot be made finite by wave-function renormalizations alone. This problem is analyzed in more detail in Appendix A, where it is shown that if a photon regulator is introduced to make the diagrams of Fig. 1 finite, then energy-momentum-tensor conservation requires the introduction of extra contributions, proportional to the mass squared of the regulator field, in the $\theta_{\mu\nu}$-regulator photon vertex. These terms may be thought of as arising from the energy-momentum tensor of the regulator field. In the limit of infinite photon regulator mass these contributions survive and, in lowest relevant order, give a second logarithmic divergence, which just cancels the logarithmic divergence of the diagrams in Fig. 1. Thus, C_1 and C_2 remain finite, and the correct form of Eq. (2.1) is actually

$$\theta_\mu{}^\mu = C_1 m_0\bar{\psi}\psi + C_2 N_0[F_{\lambda\sigma}F^{\lambda\sigma}] + \text{discontinuous terms}\,, \tag{2.2}$$

with $N_0[F_{\lambda\sigma}F^{\lambda\sigma}]$ a subtracted form of the operator $Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}$. Once it is apparent that a subtracted operator appears in Eq. (2.2), it is convenient to reexpress this operator in terms of another subtracted operator $N[F_{\lambda\sigma}F^{\lambda\sigma}]$ defined by

$$\begin{aligned}&\langle e(p)|N[F_{\lambda\sigma}F^{\lambda\sigma}]|e(p')\rangle \underset{p'\to p}{=} \langle e(p)|Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|e(p)\rangle_{\text{tree}} = 0\,,\\ &\langle 0|N[F_{\lambda\sigma}F^{\lambda\sigma}]|\gamma(p,\epsilon_1)\gamma(-p',\epsilon_2)\rangle \underset{p'\to p}{=} \langle 0|Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle_{\text{tree}}\end{aligned} \tag{2.3}$$

through a relation of the form

$$N_0[F_{\lambda\sigma}F^{\lambda\sigma}] = aN[F_{\lambda\sigma}F^{\lambda\sigma}] + bm_0\bar{\psi}\psi + \text{discontinuous term}\,. \tag{2.4}$$

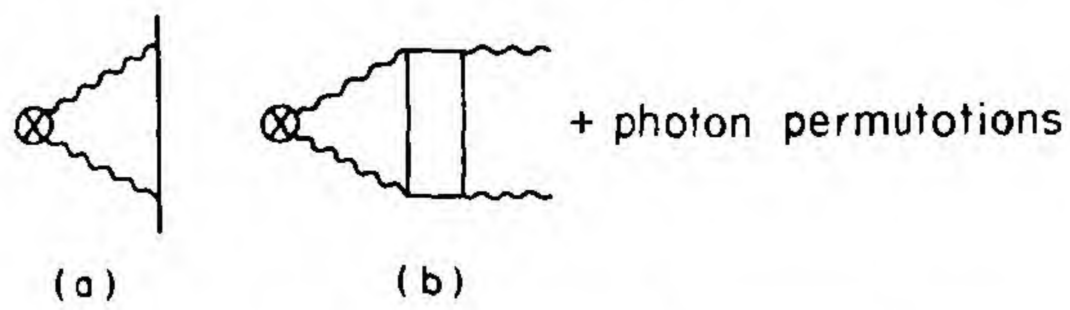


FIG. 1. (a), (b) Logarithmically divergent electron and photon vertex parts, respectively, of the operator $Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}$, the coupling of which is denoted by $\otimes$. Wavy lines indicate photon propagators, and solid lines indicate electron propagators.

This leads to the final operator form for the trace equation

$$\theta_\mu{}^\mu = K_1 m_0\bar{\psi}\psi + K_2 N[F_{\lambda\sigma}F^{\lambda\sigma}] + \text{discontinuous term}\,, \tag{2.5}$$

with the subtracted operator $N[F_{\lambda\sigma}F^{\lambda\sigma}]$ uniquely specified by the conditions of Eq. (2.3).

We proceed now to determine the coefficients K_1 and K_2 in Eq. (2.5) by taking matrix elements of Eq. (2.5) between appropriate sets of states. Taking first the matrix element between electron states in the limit of zero momentum transfer, and using[10]

$$\langle e(p)|\theta_\mu{}^\mu|e(p')\rangle \underset{p'\to p}{=} \eta^{\mu\nu}\left(\frac{p_\mu p_\nu + p_\nu p_\mu}{2m}\right) = m\,, \tag{2.6}$$

and Eq. (2.3) we find

$$K_1\langle e(p)|m_0\bar{\psi}\psi|e(p)\rangle = m\,. \tag{2.7}$$

However, as shown by Sato[11] and as explained in Appendix B, it is easy to see from the Callan-Symanzik equation for the electron propagator that

$$\langle e(p)|m_0\bar{\psi}\psi|e(p)\rangle = \frac{m}{1+\delta(\alpha)}\,, \tag{2.8}$$

with $\delta(\alpha)$ the function of the fine-structure constant α defined by[12]

$$1+\delta(\alpha) = \frac{m}{m_0}\frac{\partial m_0}{\partial m}\,. \tag{2.9}$$

Combining Eqs. (2.7) and (2.8), we conclude that[13]

$$K_1 = 1+\delta(\alpha) = 1 + \frac{3\alpha}{2\pi} + \cdots\,. \tag{2.10}$$

Next we take the matrix element of Eq. (2.2) between the vacuum and the two-photon state, again in the limit of zero momentum transfer. Now as Iwasaki[14] has shown, the general form of the vertex $\langle 0|\theta_{\mu\nu}|\gamma(p_1,\epsilon_1)\gamma(p_2,\epsilon_2)\rangle$ is

$$\begin{aligned}\langle 0|\theta_{\mu\nu}|\gamma(p_1,\epsilon_1)\gamma(p_2,\epsilon_2)\rangle &= [\tfrac{1}{2}(F^{1\rho}_{\mu}F^{2}_{\nu\rho} + F^{1\rho}_{\nu}F^{2}_{\mu\rho}) - \tfrac{1}{4}\eta_{\mu\nu}F^{1\lambda\sigma}F^{2}_{\lambda\sigma}]A(q^2)\\ &\quad + F^{1\lambda\sigma}F^{2}_{\lambda\sigma}(p_1-p_2)_\mu(p_1-p_2)_\nu B(q^2) + \tfrac{1}{2}(F^{1}_{\mu\alpha}F^{2}_{\nu\beta} + F^{1}_{\nu\alpha}F^{2}_{\mu\beta})q^\alpha q^\beta C(q^2)\,,\end{aligned} \tag{2.11}$$

$$q = p_1 + p_2,\quad p_1{}^2 = p_2{}^2 = 0,\quad F^{i}_{\alpha\beta} = (p_i)_\alpha(\epsilon_i)_\beta - (p_i)_\beta(\epsilon_i)_\alpha,\quad i = 1, 2\,.$$

As Iwasaki notes, Eq. (2.11) implies that the vacuum-to-two-photon matrix element of $\theta_\mu{}^\mu$ is

$$\langle 0|\theta_\mu{}^\mu|\gamma(p_1,\epsilon_1)\gamma(p_2,\epsilon_2)\rangle = (\epsilon_1\cdot\epsilon_2 p_1\cdot p_2 - \epsilon_1\cdot p_2\epsilon_2\cdot p_1) \times q^2[-2B(q^2)+\tfrac{1}{2}C(q^2)]\ , \quad (2.12)$$

which vanishes at $q^2=0$. Hence, from the vacuum-to-two-photon matrix element of Eq. (2.5) we get, using Eq. (2.6),

$$0=[1+\delta(\alpha)]\langle 0|m_0\bar\psi\psi|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle + K_2\langle 0|Z_3{}^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle_{\text{tree}}\ . \quad (2.13)$$

Now as shown by Adler *et al.*[15] and again as explained in Appendix B, from the Callan-Symanzik equation for the photon propagator one sees that

$$\langle 0|m_0\bar\psi\psi|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle = -\frac{1}{4}\frac{\beta(\alpha)}{1+\delta(\alpha)} \times\langle 0|Z_3{}^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle_{\text{tree}}\ , \quad (2.14)$$

with $\beta(\alpha)$ defined by[12,16]

$$\beta(\alpha)=\frac{1}{\alpha}m\frac{\partial\alpha}{\partial m} = \frac{1}{\alpha}[1+\delta(\alpha)]m_0\frac{\partial\alpha}{\partial m_0} = \frac{2\alpha}{3\pi}+\frac{\alpha^2}{2\pi^2}+\cdots\ . \quad (2.15)$$

Comparing Eq. (2.13) with Eq. (2.14), we learn that

$$K_2=\tfrac{1}{4}\beta(\alpha)\ , \quad (2.16)$$

and thus our final result for the trace equation is

$$\theta_\mu{}^\mu=[1+\delta(\alpha)]m_0\bar\psi\psi+\tfrac{1}{4}\beta(\alpha)N[F_{\lambda\sigma}F^{\lambda\sigma}] + \text{discontinuous term}\ . \quad (2.17)$$

The first two terms in the power-series expansion of the coefficient of the $F_{\lambda\sigma}F^{\lambda\sigma}$ term in Eq. (2.17) agree with the fourth-order calculation of Chanowitz and Ellis.[17]

The above derivation is evidently closely analogous to the derivation,[18] by use of the Callan-Symanzik equations, of the nonrenormalization theorem for the axial-vector divergence anomaly

$$\frac{\partial}{\partial x_\mu}j_\mu^5(x)=2im_0j^5(x) + \frac{\alpha_0}{4\pi}F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho}\ . \quad (2.18)$$

However, there are two important ways in which the trace anomaly differs from the axial-vector divergence anomaly. First, the trace anomaly *is* renormalized in higher orders of perturbation theory, and in fact would vanish, leaving only the "soft" operator $[1+\delta(\alpha)]m_0\bar\psi\psi$ as the trace, if $\beta(\alpha)$ satisfied the eigenvalue condition[12,19]

$$\beta(\alpha)=0\ . \quad (2.19)$$

Second, whereas the axial anomaly involves the *divergent* operator $Z_3{}^{-1}F^{\xi\sigma}F^{\tau\rho}\epsilon_{\xi\sigma\tau\rho}$, with the consequence that matrix elements of j_μ^5 are not renormalized by wave-function renormalization factors alone, the trace anomaly involves the *convergent* (once-subtracted) operator $N[F_{\lambda\sigma}F^{\lambda\sigma}]$, consistent with the finiteness of matrix elements of the energy-momentum tensor. The appearance of a subtracted operator in Eq. (2.17), as well as closely analogous results of Lowenstein and Schroer in ϕ^4 scalar field theory,[19] suggests that it should be natural to derive Eq. (2.17) within the framework of the normal-product formalism.[5] This is the subject to which we now turn.

III. NORMAL-PRODUCT DERIVATION

In all subsequent discussion we assume that the vacuum expectation value of any operator we consider has been implicitly subtracted off.

In this section we will express $\theta_\mu{}^\mu$ as a linear combination of normal-product operators. Underlying this derivation are the following two observations:

(1) The expression for $\theta_\mu{}^\mu$ in terms of normal products is determined entirely by its insertions at zero momentum into Green's functions: The only operators that can occur are gauge invariant and of dimension at most 4; but the only such operator which vanishes at zero momentum is $\partial^\mu(\bar\psi\gamma_\mu\psi)$, and this operator has the wrong charge-conjugation properties.

(2) The Callan-Symanzik equation is the Ward identity which expresses the nonconservation of the dilatation current[20] and the divergence of the dilation current is essentially $\theta_\mu{}^\mu$. So, if we express this Ward identity in terms of an insertion of $\int\theta_\mu{}^\mu d^4x$, then comparison with the Callan-Symanzik equation in its standard form will give $\theta_\mu{}^\mu$ (at zero momentum) in terms of renormalized operators.

We will use dimensional renormalization[21] to define both the normal products and the renormalized Green's functions. This is by no means essential: All that is required is that the subtractions performed implicitly by the normal products agree with those obtained by an explicit redefinition of the fields and parameters of the bare theory.

We will frequently consider insertions at zero momentum of operators in Green's functions. In

Lowenstein's[22] terminology these are differential vertex operations (DVO's).

First we must define the theory by adding a gauge-fixing term

$$\mathcal{L}_{gf} \equiv -\tfrac{1}{2}(\partial\cdot A)^2/\xi_0 \tag{3.1}$$

to the Lagrangian so the theory is given by

$$\mathcal{L} = \mathcal{L}_{inv} + \mathcal{L}_{gf}\ . \tag{3.2}$$

As usual ξ_0 is renormalized by writing

$$\xi_0 = Z_3\xi_R\ . \tag{3.3}$$

We are now ready to start the proof.

Consider the equation of dimensional analysis for an unrenormalized (but dimensionally regularized) Green's function G_0:

$$0=\left(\kappa\frac{\partial}{\partial\kappa}-D_G+m_0\frac{\partial}{\partial m_0}+(2-\tfrac{1}{2}n)e_0\frac{\partial}{\partial e_0}\right)G_0\ . \tag{3.4}$$

Here D_G is the mass dimension of G_0.

By the action principle we can express $\partial/\partial e_0$ and $\partial/\partial m_0$ in terms of operator insertions. Thus,

$$m_0\frac{\partial}{\partial m_0}=-im_0\bar\psi\psi^{\sim}(0)\ , \tag{3.5}$$

$$e_0\frac{\partial}{\partial e_0}=-ie_0\bar\psi A\!\!\!/\,\psi^{\sim}(0)\ , \tag{3.6}$$

where the superscript tilde means that the operator has been Fourier transformed into momentum space. Then

$$0=\left(\kappa\frac{\partial}{\partial\kappa}-D_G-i[m_0\bar\psi\psi+(2-\tfrac{1}{2}n)e_0\bar\psi A\!\!\!/\,\psi]^{\sim}(0)\right)G_R\ , \tag{3.7}$$

where we have multiplied the equation by $Z_2^{-1/2}$ for each external fermion line of G_0, and by $Z_3^{-1/2}$ for each external photon. Thus, Eq. (3.7) is an equation for the renormalized Green's function G_R.

To rewrite (3.7) in terms of $\theta_\mu{}^\mu$ we will need the counting identities.[22] These are simple consequences of the equations of motion, and can be written in terms of either bare fields or normal products. In QED these identities are

$$\begin{aligned}N_e&=(\tfrac{1}{2}\bar\psi\overleftrightarrow{D\!\!\!/}\psi+im_0\bar\psi\psi)^{\sim}(0)\\&=(\tfrac{1}{2}N[\bar\psi\overleftrightarrow{D\!\!\!/}\psi]+imN[\bar\psi\psi])^{\sim}(0)\ ,\end{aligned} \tag{3.8}$$

$$\begin{aligned}N_\gamma&=[\tfrac{1}{2}iF_{\mu\nu}{}^2+i(\partial\cdot A)^2/\xi_0+ie_0\bar\psi A\!\!\!/\,\psi]^{\sim}(0)\\&=\{\tfrac{1}{2}iN[F_{\mu\nu}{}^2]+iN[(\partial\cdot A)^2]\xi_R\\&\quad+ie\mu^{2-n/2}N[\bar\psi A\!\!\!/\,\psi]\}^{\sim}(0)\ .\end{aligned} \tag{3.9}$$

Here N_e and N_γ denote respectively the number of external electron lines of a Green's function and the number of external photon lines. Also, μ is the unit of mass,[21] which is used by dimensional regularization to make explicit the dimension of e_0, while keeping dimensionless the renormalized charge e; thus we have $e_0=\mu^{2-n/2}eZ_3(e,n)^{-1/2}$. These identities are for operations applied to Green's functions, i.e., for DVO's.

We can now write

$$\begin{aligned}\tilde\theta_\mu{}^\mu(0)&=(2-\tfrac{1}{2}n)iN_\gamma+\tfrac{1}{2}(1-n)iN_e\\&\quad+\{(2-\tfrac{1}{2}n)(\partial\cdot A)^2/\xi_0+m_0\bar\psi\psi\\&\quad+(2-\tfrac{1}{2}n)e_0\bar\psi A\!\!\!/\,\psi\}^{\sim}(0)\ .\end{aligned} \tag{3.10}$$

Notice that the right-hand side of (3.10) contains (a) the operators occurring in Eq. (3.7), (b) N_e and N_γ, which have been expressed in terms of renormalized operators, and (c) $(n-4)(\partial\cdot A)^2/\xi_0$. The only operator in an inconvenient form is the last one.

However,[23] an application of the gauge Ward identities to each $\partial\cdot A$ in turn proves that $(\partial\cdot A)^2/\xi_0$ has only a single-loop divergence, and that

$$\begin{aligned}\frac{1}{2\xi_0}(\partial\cdot A)^2&=\frac{1}{2\xi_R}N[(\partial\cdot A)^2]\\&\quad-\frac{ie^2\xi_R}{16\pi^2(n-4)}(\bar\psi\overleftrightarrow{D\!\!\!/}\psi+2im_0\bar\psi\psi)\ .\end{aligned} \tag{3.11}$$

Hence,

$$\begin{aligned}0&=\left[\kappa\frac{\partial}{\partial\kappa}-D_G-i\tilde\theta_\mu{}^\mu(0)+\left(\frac{3}{2}-\frac{e^2\xi_R}{8\pi^2}\right)N_e\right]G_R\\&\quad+O(n-4)\ .\end{aligned} \tag{3.12}$$

We have not yet proved $\theta_\mu{}^\mu$ to be finite, so we cannot set $n=4$ here.

Next, we recall the Callan-Symanzik equation[24,25] for G_R:

$$\begin{aligned}0=\Big(&\kappa\frac{\partial}{\partial\kappa}-D_G-\beta\frac{\partial}{\partial e}+(1+\gamma_m)m\frac{\partial}{\partial m}\\&+\gamma_3\xi_R\frac{\partial}{\partial\xi_R}-\tfrac{1}{2}\gamma_2N_e-\tfrac{1}{2}\gamma_3N_\gamma\Big)G_R\ .\end{aligned} \tag{3.13}$$

Comparison of the last two equations shows that $\tilde\theta_\mu{}^\mu(0)$ is finite at $n=4$, and that

$$\begin{aligned}\tilde\theta_\mu{}^\mu(0)&=-i\beta\frac{\partial}{\partial e}+i(1+\gamma_m)m\frac{\partial}{\partial m}+i\gamma_3\xi_R\frac{\partial}{\partial\xi_R}\\&\quad-i\left(\tfrac{1}{2}\gamma_2+\tfrac{3}{2}-\frac{e^2\xi_R}{16\pi^2}\right)N_e-\frac{i}{2}\gamma_3N_\gamma\\&=\Big[-\beta N[\bar\psi A\!\!\!/\,\psi]+(1+\gamma_m)m_0\bar\psi\psi-\tfrac{1}{2}\gamma_3N[(\partial\cdot A)^2]/\xi_R\\&\quad-i\left(\tfrac{1}{2}\gamma_2+\tfrac{3}{2}-\frac{e^2\xi_R}{16\pi^2}\right)N_e-\tfrac{1}{2}i\gamma_3N_\gamma\Big]^{\sim}(0)\ .\end{aligned} \tag{3.14}$$

Here the renormalized action principle has been used to express derivatives with respect to e etc. in terms of normal products. Also, we have used the result[9] that $m_0\bar\psi\psi = mN[\bar\psi\psi]$.

Finally, we use (a) the identities (3.8) and (3.9) to express N_e and N_γ in terms of normal products, (b) the result $\beta = e\gamma_3/2$, and (c) the observation made earlier that the zero-momentum expression for $\theta_\mu{}^\mu$ determines the expression at all momenta. We get[26]

$$\theta_\mu{}^\mu = \tfrac14 \gamma_3 N[F_{\mu\nu}{}^2] + (1+\gamma_m) m_0\bar\psi\psi - [\gamma_2 + 3 - e^2\xi_R/(8\pi^2)] \times (\tfrac12 iN[\bar\psi\overleftrightarrow{\not\partial}\psi] - mN[\bar\psi\psi]) \,. \tag{3.15}$$

Use of the fermion equations of motion gives[27]

$$\theta_\mu{}^\mu = \tfrac14 \gamma_3 N[F_{\mu\nu}{}^2] + (1+\gamma_m) m_0\bar\psi\psi \,, \tag{3.16}$$

the same operator formula as was found in Eq. (2.17) above.

Note added in proof. After this work was completed, we learned that essentially identical results have been obtained by N. K. Nielsen (unpublished).

ACKNOWLEDGMENTS

We wish to thank E. Eichten, D. Z. Freedman, and E. Weinberg for useful discussions, and M. A. B. Bég for pointing out an error in the original version of this paper and a number of helpful conversations. Two of us (S.L.A. and A.D.) wish to acknowledge the hospitality of the Aspen Center for Physics, where part of this work was done, and J.C.C. and A.D. acknowledge the hospitality of the Stanford Linear Accelerator Center.

APPENDIX A

We analyze here the consequences of including a photon regulator to make finite the divergent diagrams of Figs. 1(a) and 1(b). It proves convenient to use a regulator scheme similar to that used[28] in studying the axial-vector divergence anomaly, and specified as follows:

(i) The smallest closed fermion loops illustrated in Fig. 2(a) are given their usual gauge-invariant, renormalized values.

(ii) The larger fermion loops, such as illustrated in Fig. 2(b), are calculated to be photon gauge-invariant and hence finite.

(iii) All photon propagators are regularized: Photon propagators emerging singly from vertices, as in Fig. 2(c), are regularized by the replacement

$$\frac{1}{p^2} \to \frac{1}{p^2} - \frac{1}{p^2 - M^2} = \frac{-M^2}{p^2(p^2-M^2)} \,, \tag{A1}$$

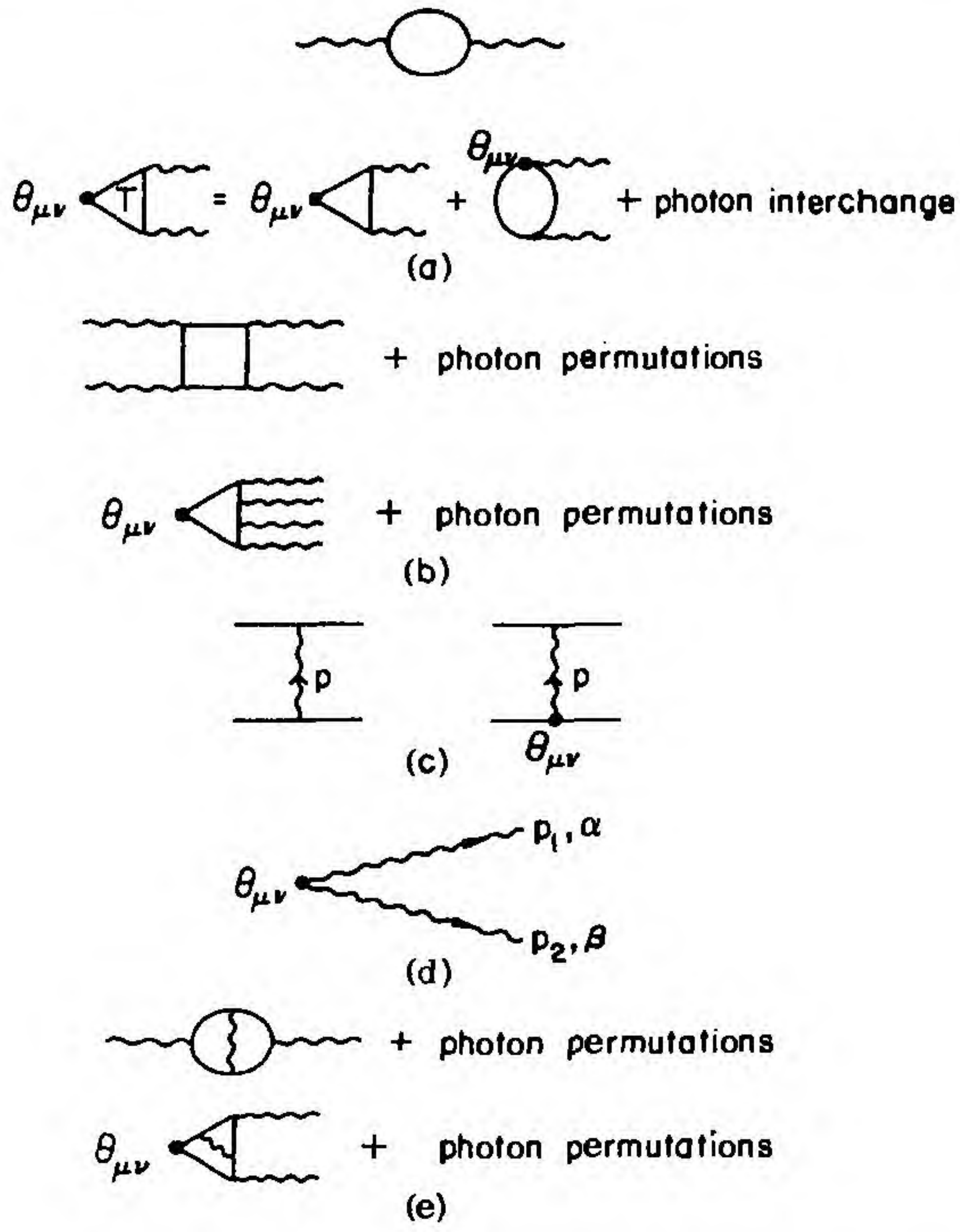


FIG. 2. (a) Smallest closed fermion loops which are given their gauge-invariant fully renormalized values. (b) Larger fermion loops which are evaluated to be gauge invariant. (c) Photons emerging from single-photon vertices, which are regulated according to Eq. (A1). (d) Photon pair emerging from $\theta_{\mu\nu}$, which is regulated according to Eq. (A2). (e) Fermion-loop diagrams with photon radiative corrections.

with M the regulator mass. Pairs of photon propagators emerging from the energy-momentum tensor $\theta_{\mu\nu}$, as in Fig. 2(d), are regularized by the replacement

$$\frac{1}{p_1{}^2} V_{\mu\nu\alpha\beta}(p_1,p_2)\frac{1}{p_2{}^2} \to \frac{1}{p_1{}^2} V_{\mu\nu\alpha\beta}(p_1,p_2)\frac{1}{p_2{}^2} - \frac{1}{p_1{}^2 - M^2} V^M_{\mu\nu\alpha\beta}(p_1,p_2)\frac{1}{p_2{}^2 - M^2} \,, \tag{A2}$$

with the regulator vertex $V^M_{\mu\nu\alpha\beta}$ chosen so that (apart from photon gauge terms, which do not contribute to on-shell matrix elements) the algebraic structure of the gravitational Ward identities implied by conservation of $\theta_{\mu\nu}$ is preserved. Specifically, the Feynman rules for vertices of $\theta_{\mu\nu}$ give

$$V_{\mu\nu\alpha\beta}(p_1,p_2) = -\tfrac{1}{2}\eta_{\mu\nu}(p_1\cdot p_2\eta_{\alpha\beta} - p_{1\beta}p_{2\alpha}) + \tfrac{1}{2}(p_1\cdot p_2\eta_{\mu\alpha}\eta_{\nu\beta} + p_{1\mu}p_{2\nu}\eta_{\alpha\beta} - p_{1\mu}p_{2\alpha}\eta_{\nu\beta} - p_{2\nu}p_{1\beta}\eta_{\mu\alpha}$$
$$+ p_1\cdot p_2\eta_{\nu\alpha}\eta_{\mu\beta} + p_{1\nu}p_{2\mu}\eta_{\alpha\beta} - p_{1\nu}p_{2\alpha}\eta_{\mu\beta} - p_{2\mu}p_{1\beta}\eta_{\nu\alpha})\,, \tag{A3}$$

which when contracted with $(p_1+p_2)^\mu$ gives

$$(p_1+p_2)^\mu V_{\mu\nu\alpha\beta}(p_1,p_2) = \text{gauge terms} + \tfrac{1}{2}p_1{}^2(p_{2\nu}\eta_{\alpha\beta} - p_{2\alpha}\eta_{\nu\beta}) + \tfrac{1}{2}p_2{}^2(p_{1\nu}\eta_{\alpha\beta} - p_{1\beta}\eta_{\nu\alpha})\,. \tag{A4}$$

We wish to construct $V^M_{\mu\nu\alpha\beta}(p_1,p_2)$ so that

$$(p_1+p_2)^\mu V^M_{\mu\nu\alpha\beta}(p_1,p_2) = \text{gauge terms} + \tfrac{1}{2}(p_1{}^2 - M^2)(p_{2\nu}\eta_{\alpha\beta} - p_{2\alpha}\eta_{\nu\beta}) + \tfrac{1}{2}(p_2{}^2 - M^2)(p_{1\nu}\eta_{\alpha\beta} - p_{1\beta}\eta_{\nu\alpha})\,, \tag{A5}$$

which gives for the divergence of Eq. (A2)

$$(p_1+p_2)^\mu\left(\frac{1}{p_1{}^2}V_{\mu\nu\alpha\beta}(p_1,p_2)\frac{1}{p_2{}^2} - \frac{1}{p_1{}^2 - M^2}V^M_{\mu\nu\alpha\beta}(p_1,p_2)\frac{1}{p_2{}^2 - M^2}\right)$$
$$= \text{gauge terms} + \tfrac{1}{2}(p_{2\nu}\eta_{\alpha\beta} - p_{2\alpha}\eta_{\nu\beta})\left(\frac{1}{p_2{}^2} - \frac{1}{p_2{}^2 - M^2}\right) + \tfrac{1}{2}(p_{1\nu}\eta_{\alpha\beta} - p_{1\beta}\eta_{\nu\alpha})\left(\frac{1}{p_1{}^2} - \frac{1}{p_1{}^2 - M^2}\right), \tag{A6}$$

which has the same structure as the divergence of $(1/p_1{}^2)V_{\mu\nu\alpha\beta}(1/p_2{}^2)$, apart from the replacement of the photon propagators by regularized propagators. One easily finds that the lowest-order polynomial in momenta satisfying Eq. (A5) is

$$V^M_{\mu\nu\alpha\beta}(p_1,p_2) = V_{\mu\nu\alpha\beta}(p_1,p_2) - \tfrac{1}{2}M^2(\eta_{\nu\mu}\eta_{\alpha\beta} - \eta_{\mu\alpha}\eta_{\nu\beta} - \eta_{\mu\beta}\eta_{\nu\alpha})\,. \tag{A7}$$

Thus, the requirement that the regularization scheme respect gravitational Ward identities introduces an explicit M^2 dependence into the $\theta_{\mu\nu}$-photon vertex. This is, of course, just the contribution to $\theta_{\mu\nu}$ expected from the mass term in the regulator field Lagrangian.

(iv) The regularization prescription adopted above makes radiative correction diagrams such as illustrated in Fig. 2(e) finite for finite M, but divergent as $M\to\infty$, with the divergences canceled by appropriate counterterms appearing in the renormalization constant $Z_3(M)$. We note, however, that since explicitly renormalized values for the single-loop diagrams of Fig. 2(a) are always used, Z_3 contains no counterterms referring to these diagrams. In effect, we have adopted a type of intermediate renormalization procedure, in which Z_3 contains counterterms only for those vacuum polarization graphs which involve internal virtual photons.

Having specified the regularization procedure, we can now turn to a study of the lowest-order divergent $\theta_\mu{}^\mu$ insertions of Fig. 1. It suffices to consider these insertions at zero four-momentum transfer, since the *difference* between zero and nonzero four-momentum transfer will converge. Focussing on the trace-to-two-photon vertex on the left-hand side of the dashed line in Fig. 3(a), we find in one-fermion-loop order that there are two classes of $\theta_{\mu\nu}$ couplings which contribute, as illustrated in Fig. 2(b) and Fig. 2(c). [We note in passing that an explicit check of $\theta_{\mu\nu}$ conservation for the diagrams of Figs. 2(b) and 2(c) shows that the structure of the Ward identities is guaranteed by the regularization scheme sketched above, with no need for any additional vertex modifications beyond that given by Eq. (A7).] Taking the trace on $\mu\nu$ of Fig. 3(b), using the trace anomaly formula of Eq. (2.17) to leading order, and dropping gauge terms gives

$$\eta^{\mu\nu}[3(b)] = \frac{-2\alpha}{3\pi}\eta_{\alpha\beta}\frac{M^4}{p^2(p^2-M^2)^2}\,. \tag{A8}$$

In the absence of regulators, the trace of Fig. 3(c) would vanish, but when regulators are included it is nonvanishing, on account of the term propor-

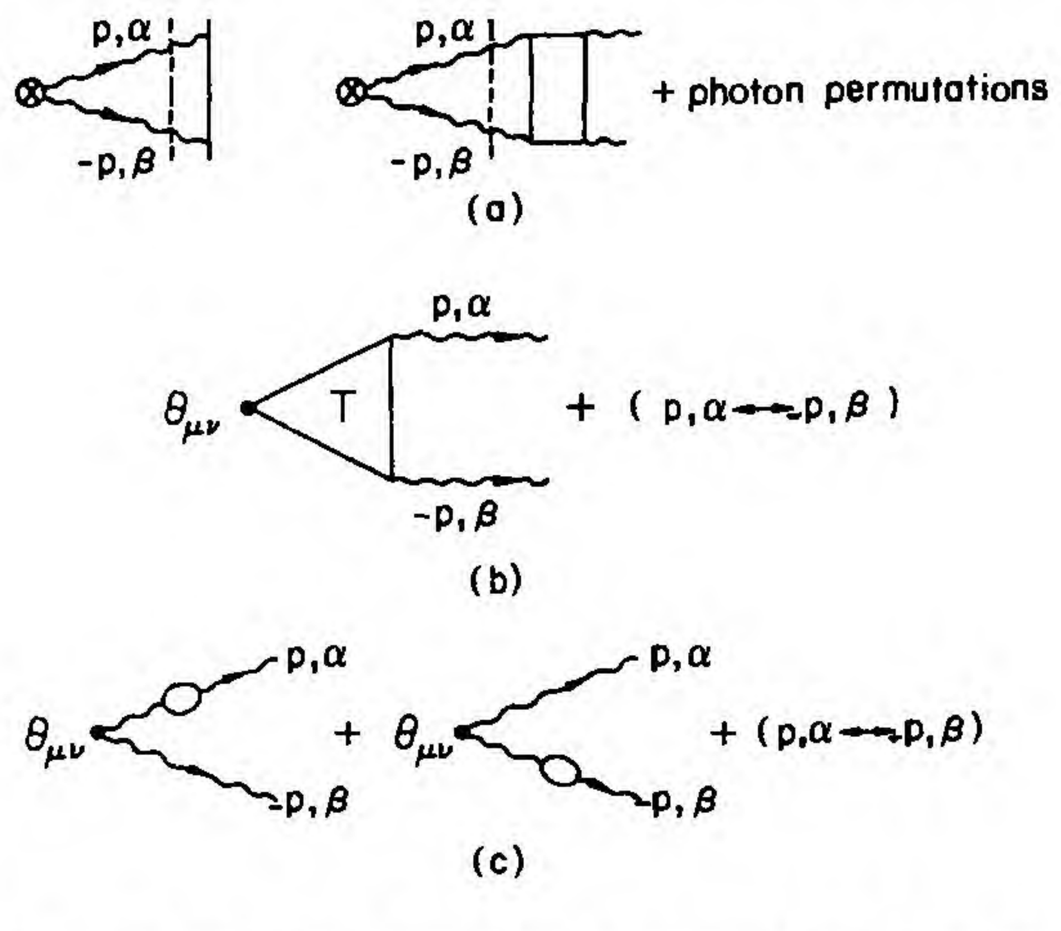


FIG. 3. (a) The divergent diagrams of Fig. 1, at zero four-momentum transfer. We focus on the trace-to-two-photon vertex on the left-hand side of the dashed line. (b), (c) Classes of one-fermion-loop diagrams which contribute to the left-hand side of the dashed line in (a).

tional to M^2 in Eq. (A7), and one finds

$$\eta^{\mu\nu}[3(c)] = \frac{-4M^4\eta_{\alpha\beta}\bar{\Pi}^{(2)}(p^2/m^2,\alpha)}{(p^2-M^2)^3}. \quad \text{(A9)}$$

Setting $-p^2 = x$, and using the fact that

$$\bar{\Pi}^{(2)}(p^2/m^2,\alpha) \underset{x\to\infty}{\simeq} -\frac{\alpha}{3\pi}\ln x - c, \quad \text{(A10)}$$

with c a constant, the sum of Eqs. (A8) and (A9) becomes

$$\begin{aligned}\eta^{\mu\nu}[3(b)] + \eta^{\mu\nu}[3(c)] &= \frac{2\alpha}{3\pi}\eta_{\alpha\beta}\frac{M^4}{x(x+M^2)^2} - \frac{4\eta_{\alpha\beta}M^4(\alpha/3\pi\ln x + c)}{(x+M^2)^3}\\ &= 2\eta_{\alpha\beta}M^4\frac{d}{dx}\left(\frac{\alpha/3\pi\ln x + c}{(x+M^2)^2}\right). \quad \text{(A11)}\end{aligned}$$

Now the leading single logarithmic divergence of either of the diagrams in Fig. 3(a) comes from an integral of the form

$$\int^{\infty} x dx\{\eta^{\mu\nu}[3(b)] + \eta^{\mu\nu}[3(c)]\}\psi(x), \quad \text{(A12)}$$

where $\psi(x) \sim c_1/x + \cdots$ represents the right-hand side of the dashed line. But substituting Eq. (A11) and the leading term of $\psi(x)$ into Eq. (A12), we get a result proportional to

$$M^4\int^{\infty} dx\frac{d}{dx}\left(\frac{\alpha/3\pi\ln x + c}{(x+M^2)^2}\right) = M^4\left(\frac{\alpha/3\pi\ln x + c}{(x+M^2)^2}\right)\Bigg|_{x\ \text{finite}}^{x=\infty}, \quad \text{(A13)}$$

which approaches a finite limit as the regulator mass M approaches infinity. In other words, the logarithmically divergent contributions to the trace coming from Figs. 3(b) and Figs. 3(c) precisely cancel: in effect, the extra M^2 term in the $\theta_{\mu\nu}$-regulator photon vertex of Eq. (A7) generates, in the limit as $M\to\infty$, a subtraction counterterm for the divergent operator $Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}$. The mechanism operating here is evidently a photon analog of the fermion regulator behavior[3] which can be thought of as producing the trace anomaly in the first place.

APPENDIX B

We give here the derivation of Eqs. (2.8) and (2.14), and also illustrate Iwasaki's theorem on the vanishing of $\langle 0|\theta_\mu{}^\mu|\gamma\gamma\rangle$ in a special case. To derive Eq. (2.8) we follow the method of Sato.[11] Introducing the scalar vertex part $\bar{\Gamma}_s(p_1,p_2)$, we have

$$\langle e(p)|m_0\bar{\psi}\psi|e(p)\rangle = \bar{\Gamma}_s(p,p)|_{\not p = m}. \quad \text{(B1)}$$

Writing

$$\begin{aligned}\bar{\Gamma}_s(p,p) &= -iZ_2 m_0\frac{\partial}{\partial m_0}[S_F'(p)]^{-1}\\ &= -Z_2 m_0\frac{\partial}{\partial m_0}Z_2^{-1}[\not p - m - \bar{\Sigma}(p)], \quad \text{(B2)}\end{aligned}$$

with S_F' the unrenormalized electron propagator and $\bar{\Sigma}$ the renormalized electron proper self-energy, and substituting into Eq. (B1), we get

$$\bar{\Gamma}_s(p,p)|_{\not p = m} = m_0\frac{\partial m}{\partial m_0} + \left(m_0\frac{\partial}{\partial m_0}\bar{\Sigma}(p)\right)\Bigg|_{\not p = m}. \quad \text{(B3)}$$

Now by the chain rule we have

$$\left(m_0\frac{\partial}{\partial m_0}\bar{\Sigma}(p)\right)\Bigg|_{\not p = m} = \left(m_0\frac{\partial m}{\partial m_0}\frac{\partial\bar{\Sigma}}{\partial m} + m_0\frac{\partial\alpha}{\partial m_0}\frac{\partial\bar{\Sigma}}{\partial\alpha}\right)_{\not p = m}, \quad \text{(B4)}$$

while from the fact that $\bar{\Sigma}$ is homogeneous of degree 1 in $\not p$ and m we get

$$\bar{\Sigma} = \left(m\frac{\partial}{\partial m} + \not p\frac{\partial}{\partial\not p}\right)\bar{\Sigma}. \quad \text{(B5)}$$

Combining Eqs. (B4) and (B5), we see that the renormalization conditions on $\bar{\Sigma}$,

$$\bar{\Sigma}|_{\not p = m} = \frac{\partial\bar{\Sigma}}{\partial\not p}\Bigg|_{\not p = m} = 0 \quad \text{(B6)}$$

imply that

$$\left(m_0\frac{\partial}{\partial m_0}\bar{\Sigma}(p)\right)\Bigg|_{\not p = m} = 0. \quad \text{(B7)}$$

[In Eq. (B6) we have assumed the Yennie gauge, in which $\bar{S}_F'(\not p)$ has a true pole at $\not p = m$; this restriction is immaterial, since the final result of Eq. (B8) is manifestly gauge invariant.] Thus, the second term in Eq. (B3) vanishes, giving the desired result

$$\langle e(p)|m_0\bar{\psi}\psi|e(p)\rangle = m_0\frac{\partial m}{\partial m_0} = \frac{m}{1+\delta(\alpha)}. \quad \text{(B8)}$$

To derive Eq. (2.14) we follow a similar procedure. Introducing the zero-momentum-transfer scalar to two-photon vertex $\bar{\Gamma}_{\gamma\gamma s}(p^2/m^2,\alpha)$, we have

$$\begin{aligned}&\langle 0|m_0\bar{\psi}\psi|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle\\ &= \tfrac{1}{4}\alpha\bar{\Gamma}_{\gamma\gamma s}(0,\alpha)\\ &\quad\times\langle 0|Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle_{\text{tree}}. \quad \text{(B9)}\end{aligned}$$

However, $\bar{\Gamma}_{\gamma\gamma s}(p^2/m^2,\alpha)$ is related to the photon renormalized proper self-energy $\bar{\Pi}(p^2/m^2,\alpha)$ by the Callan-Symanzik equation[12]

$$\frac{1}{1+\delta(\alpha)}\left(m\frac{\partial}{\partial m}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}\right)\frac{1}{\alpha}[1+\bar{\Pi}(p^2/m^2,\alpha)] = \bar{\Gamma}_{\gamma\gamma s}(p^2/m^2,\alpha)\ . \quad \text{(B10)}$$

On setting $p^2=0$ in Eq. (B10) and using the renormalization condition

$$\bar{\Pi}(0,\alpha)=0\ , \quad \text{(B11)}$$

we get

$$\alpha\bar{\Gamma}_{\gamma\gamma s}(0,\alpha)=-\frac{\beta(\alpha)}{1+\delta(\alpha)}\ , \quad \text{(B12)}$$

which when substituted into Eq. (B9) gives the desired relation

$$\langle 0|m_0\bar{\psi}\psi|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle = -\frac{1}{4}\frac{\beta(\alpha)}{1+\delta(\alpha)} \times\langle 0|Z_3^{-1}F_{\lambda\sigma}F^{\lambda\sigma}|\gamma(p,\epsilon_1)\gamma(-p,\epsilon_2)\rangle_{\text{tree}}\ . \quad \text{(B13)}$$

It is also instructive to rearrange Eq. (B10) into a slightly different form by writing

$$\Pi^{\mu\nu}(p,-p)=(p^\mu p^\nu-p^2\eta^{\mu\nu})\bar{\Pi}(p^2/m^2,\alpha)\ ,$$
$$\Delta^{\mu\nu}(p,-p)=(p^\mu p^\nu-p^2\eta^{\mu\nu})\alpha\bar{\Gamma}_{\gamma\gamma s}(p^2/m^2,\alpha)\ , \quad \text{(B14)}$$

giving

$$\Delta^{\mu\nu}(p,-p)=\frac{1}{1+\delta(\alpha)}\left[\left(2-p\cdot\frac{\partial}{\partial p}\right)+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right] \times\pi^{\mu\nu}(p,-p)-\frac{\beta(\alpha)}{1+\delta(\alpha)}(p^\mu p^\nu-p^2\eta^{\mu\nu})\ . \quad \text{(B15a)}$$

An equivalent form, suggested by Eq. (2.17), is

$$[1+\delta(\alpha)]\Delta^{\mu\nu}(p,-p)+\beta(\alpha)(p^\mu p^\nu-p^2\eta^{\mu\nu}) =\left[2-p\cdot\frac{\partial}{\partial p}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]\Pi^{\mu\nu}(p,-p)\ . \quad \text{(B15b)}$$

Equation (B15) is an exact expression, at zero momentum transfer, for the vacuum-to-two-photon matrix element of the "naive" or canonical trace $m_0\bar{\psi}\psi$.[29] Substituting the second-order perturbation formula

$$\bar{\Pi}^{(2)}(p^2/m^2,\alpha) =\frac{-2\alpha}{\pi}\int_0^1 dx\,x(1-x)\ln\left(1-\frac{p^2x(1-x)}{m^2}\right) \quad \text{(B16)}$$

into Eqs. (B10) or (B16), we recover the usual second-order perturbation theory formula

$$\alpha\bar{\Gamma}^{(2)}_{\gamma\gamma s}(p^2/m^2,\alpha) =-\frac{4\alpha}{\pi}\int_0^1 dx\,x(1-x)\frac{m^2}{m^2-p^2x(1-x)}\ . \quad \text{(B17)}$$

As a simple, explicit check on Iwasaki's theorem we have calculated the second-order vacuum-to-two-photon matrix element of $\theta_{\mu\nu}$ at zero momentum transfer. (This can either be done directly by diagrammatic techniques, or more simply by using the Ward identities[30] following from conservation of $\theta_{\mu\nu}$.) Denoting this matrix element by $T^{(2)}_{\mu\nu\alpha\beta}(p)$, with p, $-p$ the (virtual) photon four-momenta and with α and β the photon polarization indices, we find that

$$T^{(2)}_{\mu\nu\alpha\beta}(p)=t_{\mu\nu\alpha\beta}(p)\bar{\Pi}^{(2)}(p^2/m^2,\alpha) +(p_\alpha p_\beta-p^2\eta_{\alpha\beta})p_\mu p_\nu 2\frac{\partial}{\partial p^2}\bar{\Pi}^{(2)}(p^2/m^2,\alpha)\ ,$$
$$t_{\mu\nu\alpha\beta}=p^2(\eta_{\mu\nu}\eta_{\alpha\beta}-\eta_{\mu\alpha}\eta_{\nu\beta}-\eta_{\nu\alpha}\eta_{\mu\beta}) -2\eta_{\alpha\beta}p_\mu p_\nu-\eta_{\mu\nu}p_\alpha p_\beta+\eta_{\mu\alpha}p_\nu p_\beta +\eta_{\nu\alpha}p_\mu p_\beta+\eta_{\nu\beta}p_\mu p_\alpha+\eta_{\mu\beta}p_\nu p_\alpha\ . \quad \text{(B18)}$$

Taking the trace we obtain

$$\eta^{\mu\nu}T^{(2)}_{\mu\nu\alpha\beta}(p)=(p_\alpha p_\beta-p^2\eta_{\alpha\beta})2p^2 \times\frac{\partial}{\partial p^2}\bar{\Pi}^{(2)}(p^2/m^2,\alpha)\ , \quad \text{(B19)}$$

which evidently vanishes for on-shell photons ($p^2=0$) as asserted by Iwasaki's general argument. To exhibit the splitting of Eq. (B19) into "naive" and anomalous trace terms, we substitute Eq. (B16) and rearrange by comparison with Eq. (B17), giving

$$\eta^{\mu\nu}T^{(2)}_{\mu\nu\alpha\beta}(p)=-(p_\alpha p_\beta-p^2\eta_{\alpha\beta}) \times\left(\alpha\bar{\Gamma}^{(2)}_{\gamma\gamma s}(p^2/m^2,\alpha)+\frac{2\alpha}{3\pi}\right) \quad \text{(B20)}$$

as expected from Eqs. (2.17) and (B9) in second order.

---

*Research sponsored by the Energy Research and Development Administration under Grant No. E(11-1)-2220.

†Research supported in part by the National Science Foundation under Grant No. MPS75-22514.

[1]We follow throughout the metric and γ-matrix conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).

[2]For reviews, see S. L. Adler, in *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser, M. Grisaru, and H. Pendleton (M.I.T. Press, Cambridge, Mass., 1970), p. 3; S. B. Treiman,

R. Jackiw, and D. J. Gross, *Lectures on Current Algebra and Its Applications* (Princeton Univ. Press, Princeton, N.J., 1972), p. 97.

[3]R. J. Crewther, Phys. Rev. Lett. 28, 1421 (1972); M. S. Chanowitz and J. Ellis, Phys. Lett. 40B, 397 (1972); M. S. Chanowitz and J. Ellis, Phys. Rev. D 7, 2490 (1973).

[4]We use e_0 and m_0 to denote the unrenormalized charge and mass, e and m to denote the renormalized charge and mass, and ψ, A_μ, and $F_{\mu\nu}$ to denote unrenormalized fields. In Sec. III, the notation for the functions defined through the Callan-Symanzik equation follows J. C. Collins and A. J. Macfarlane, Phys. Rev. D 10, 1201 (1974); the relation to the notation of Sec. II and the appendixes is $\gamma_3 = \beta$, $\gamma_m = \delta$.

[5]The normal-product algorithm was first set up using Bogoliubov-Parasiuk-Hepp-Zimmermann renormalization by W. Zimmermann, in *Lectures on Elementary Particles and Quantum Field Theory,* edited by S. Deser, M. Grisaru, and H. Pendleton (M.I.T. Press, Cambridge, Mass., 1970), Vol. I, p. 397. In Sec. III we shall use normal products defined by dimensional renormalization: J. C. Collins, Nucl. Phys. B92, 477 (1975), and P. Breitenlohner and D. Maison, Max-Planck Institut Report No. MPI-PAE PTh25 and NYU Report No. NYU/TR2/76 (unpublished).

[6]G. 't Hooft and M. Veltman, Nucl. Phys. B44, 189 (1972); C. G. Bollini and J. J. Giambiagi, Phys. Lett. 40B, 566 (1972); G. M. Cicuta and E. Montaldi, Lett. Nuovo Cimento 4, 329 (1972).

[7]C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) 59, 42 (1970). The proof of finiteness of $\theta_{\mu\nu}$ by these authors uses the trace identity to prove $\theta_\mu{}^\mu$ finite, but they ignore the trace anomaly. Our completion of their proof will be to show that the trace anomaly is finite, and hence that its existence does not affect the proof that $\theta_{\mu\nu}$ is finite.

[8]D. Z. Freedman, I. J. Muzinich, and E. J. Weinberg, Ann. Phys. (N.Y.) 87, 95 (1974); D. Z. Freedman and E. J. Weinberg, *ibid.* 87, 354 (1974). The proof of finiteness of $\theta_{\mu\nu}$ given in these papers does not include the possibility of fermion fields, and it depends on the Ward identities for the conservation of $\theta_{\mu\nu}$. Since these Ward identities are more complicated in the presence of fermions, the proof appears not to have a trivial generalization. In fact, use of the trace identity seems necessary.

[9]Finiteness of $m_0\bar\psi\psi$ is proved, for instance, by S. Adler and W. A. Bardeen, Phys. Rev. D 4, 3045 (1971); 6, 734(E) (1972). Within the framework of dimensional regularization it is also a special case of a general theorem by J. C. Collins, Nucl. Phys. B80, 341 (1974). Finiteness of $\theta_{\mu\nu}$ is a conjecture which we prove in Sec. III. The results in Sec. III depend in no way on Sec. II.

[10]We use throughout an invariant-state normalization convention. One of us (S.L.A.) wishes to thank M. A. B. Bég for a critical comment on the original version of this paper, which suggested the importance of studying the electron-to-electron, as well as the vacuum-to two-photon, matrix element of the trace equation in the heuristic derivation of Sec. II.

[11]H. Sato, University of Minnesota report (unpublished).

[12]S. L. Adler, Phys. Rev. D 5, 3021 (1972), Eq. (41).

[13]For the δ expansion, see Adler and Bardeen, Ref. 9, Eqs. (15b) and (47), where $\alpha(x)$ is used to denote the quantity called $\delta(x)$ in this paper.

[14]Y. Iwasaki, Phys. Rev. D 15, 1172 (1977).

[15]S. L. Adler, J. Lieberman, Y. J. Ng, and H-S. Tsao, Phys. Rev. D 14, 359 (1976), Appendix A, Eqs. (A30) and (A31). The matrix element corresponding to the effective Lagrangian $\mathcal{L}_{\rm kin}$ of Eq. (A31) is $i\mathcal{L}_{\rm kin}$, and the $n=1$ term of this is the vacuum-to-two-photon matrix element of $-im_0\bar\psi\psi$.

[16]For the β expansion, see for example, J. D. Bjorken and S. D. Drell, Ref. 1, Eq. (19.160).

[17]M. S. Chanowitz and J. Ellis, Phys. Rev. D 7, 2490 (1973).

[18]A. Zee, Phys. Rev. Lett. 29, 1198 (1972); J. Lowenstein and B. Schroer, Phys. Rev. D 7, 1929 (1973).

[19]Related observations in the ϕ^4 scalar field theory have been given by J. H. Lowenstein, Phys. Rev. D 4, 2281 (1971), and B. Schroer, Lett. Nuovo Cimento 2, 867 (1971). It would be interesting to see whether requiring the softness of the energy-momentum-tensor trace in non-Abelian gauge theories gives nontrivial coupling-constant eigenvalue conditions. The extension of the methods of this paper to quantum chromodynamics has been studied, and the results will be described elsewhere.

[20]S. Coleman and R. Jackiw, Ann. Phys. (N.Y.) 67, 552 (1971).

[21]Collins, and Breitenlohner and Maison, Ref. 5.

[22]J. H. Lowenstein, Commun. Math. Phys. 24, 1 (1971).

[23]J. C. Collins, Ph.D. thesis, Cambridge University, 1975 (unpublished). The argument is a dimensionally regularized version of that given by L. D. Landau and I. M. Khalatnikov, Zh. Eksp. Teor. Fiz. 29, 89 (1955) [Sov. Phys.-JETP 2, 69 (1956)]; K. Johnson and B. Zumino, Phys. Rev. Lett. 3, 351 (1959); B. Zumino, J. Math. Phys. 1, 1 (1960). A recent paper using dimensional regularization is B. Lautrup, Nucl. Phys. B105, 23 (1976). See also J. H. Lowenstein and B. Schroer, Phys. Rev. D 6, 1553 (1972).

[24]This equation, in the form we use it, is due to S. Weinberg, Phys. Rev. D 8, 3497 (1973). Note that β as used here differs from the β in Sec. II. The coefficients β, γ_2, γ_3, and γ_m are precisely those defined by Collins and Macfarlane (Ref. 4).

[25]We deliberately call Eq. (3.13) the Callan-Symanzik equation rather than the renormalization-group equation. It is true that *with dimensional renormalization* the two equations are essentially identical, for, by Refs. 4, 5, and 9, $m\partial/\partial m = m_0\partial/\partial m_0 = m_0\bar\psi\psi^\sim(0)$. Our derivation in this section of the trace anomaly is so arranged that it has a simple generalization to any renormalization prescription, provided that the normal products used are those natural to the prescription. (E.g., in the case of renormalization on-shell, we must define $N[F_{\mu\nu}{}^2]$ by Eq. (2.3).) But, in general (see Ref. 22), the two equations differ, and it is the Callan-Symanzik equation we must use as (3.13). This is because it is the Callan-Symanzik equation that contains the mass term $m_0\bar\psi\psi$, and we want to write $\theta_\mu{}^\mu$ as $m_0\bar\psi\psi$ plus an anomaly.

[26]Note that by Ref. 23, $\gamma_2 - e^2\xi_R/(8\pi^2)$ is precisely γ_2 evaluated in the Landau gauge $\xi_R = 0$.

[27]We could not use the equations of motion earlier because we were considering Green's functions containing $\theta_\mu{}^\mu$ at zero momentum.

[28]S. L. Adler and W. A. Bardeen, Phys. Rev. 182, 1517 (1969).

[29]Equation (B15) disagrees, by an over-all factor of $[1+\delta(\alpha)]^{-1}$ and the presence of the $\beta(\alpha)(\alpha\partial/\partial\alpha-1)$ term, with the fourth-order result given in the Appendix of Chanowitz and Ellis, Ref. 17, who have incorrectly generalized to fourth order the lowest-order canonical trace identity.

[30]Freedman, Muzinich, and Weinberg, Ref. 8, Eq. (3.7).

VOLUME 25, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1970

PHOTON SPLITTING IN A STRONG MAGNETIC FIELD

S. L. Adler, J. N. Bahcall,* C. G. Callan, and M. N. Rosenbluth
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 6 August 1970)

We determine the absorption coefficient and polarization selection rules for photon splitting in a strong magnetic field, and describe the possible application of our results to pulsars.

Recent work on pulsars suggests the presence of trapped magnetic fields within an order of magnitude in either direction of the electrodynamic critical field $B_{cr} = m^2/e = 4.41 \times 10^{13}$ G.[1] (Here m and e are, respectively, the electronic mass and charge.) In such intense fields, electrodynamic processes which are unobservable in the laboratory can become important. One such process, for photons with energy $\omega > 2m$, is photopair production, for which both the photon absorption coefficient and the corresponding vacuum dispersion have been calculated by Toll.[2] For $\omega < 2m$ the photopair process is kinematically forbidden, and the only[3] photon absorption

VOLUME 25, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1970

mechanism which does not require the presence of matter is photon splitting, i.e.,

$$\gamma(k) + \text{external magnetic field} \to \gamma(k_1) + \gamma(k_2). \quad (1)$$

We present in this note the results of calculations of the absorption coefficient and the polarization selection rules for this reaction, in the case of a constant and spatially uniform external magnetic field $\bar{B}$.[4]

To begin, let us consider photon splitting when dispersive effects caused by the external field are neglected, so that the photon four-momenta satisfy the vacuum dispersion relation

$$k^2 = k_1^2 = k_2^2 = 0. \quad (2)$$

Because the external field $\bar{B}$ is constant and spatially uniform, it cannot transfer four-momentum to the photons, and so the four-vectors k, k_1, k_2 must satisfy four-momentum conservation by themselves,

$$k = \omega(1, \hat{k}) = k_1 + k_2 = \omega_1(1, \hat{k}_1) + \omega_2(1, \hat{k}_2). \quad (3)$$

It is easily seen that Eqs. (2) and (3) can be satisfied only if the three propagation directions $\hat{k}$, $\hat{k}_1$, and $\hat{k}_2$ are identical, which implies that the photon four-vectors are proportional,

$$k_1 = (\omega_1/\omega)k, \quad k_2 = (\omega_2/\omega)k. \quad (4)$$

We will use Eq. (4) to simplify considerably the matrix elements for photon splitting. To leading order in e, the matrix element involving $2n+1$ interactions with the external field comes from the ring diagrams with $2n+4$ vertices which are illustrated in Fig. 1. When all permutations of the vertices are summed over, the matrix element is gauge-invariant, and therefore must couple the three photons and the external field only through their respective field-strength tensors $F_{\mu\nu}$, $F_{\mu\nu}^1$, $F_{\mu\nu}^2$, and $\bar{F}_{\mu\nu}$. Because Eq. (4) tells us that only one four-momentum is present in the problem, the matrix element for Fig. 1 is a sum of terms of the form

$$FF^1F^2 \times \underbrace{\bar{F}\cdots\bar{F}}_{2n+1 \text{ factors}} \times \underbrace{k\cdots k}_{2l \text{ factors}} \quad (5)$$

with the Lorentz indices contracted to form a Lorentz scalar (which is why the number of factors k must be even). Since $k^\mu F_{\mu\nu} = k^\mu F_{\mu\nu}^1 = k^\mu F_{\mu\nu}^2 = k^\mu \bar{F}_{\mu\nu} k^\nu = 0$, a nonvanishing contribution is obtained only if each factor k is contracted with a different $\bar{F}$, which means that we must have $l \leq n$. We will now show further that when $l = n$, the contribution to the matrix element still vanishes. Writing $v_\mu = \bar{F}_{\mu\nu}k^\nu$, a term with $2n$ factors k has the form

$$FF^1F^2\bar{F} \times \underbrace{v\cdots v}_{2n \text{ factors}}, \quad (6)$$

again with Lorentz indices contracted to form a Lorentz scalar. Because of the antisymmetry of the field strength, the number of factors v which can be contracted with field strengths can only be 0, 2, or 4. An enumeration of these contractions[5] shows that they must always contain at least one factor of the following five types: $F_{\alpha\beta}{}^1F^{2\,\alpha\beta}$, $v_\alpha F^{1\,\alpha\beta}F_{\beta\gamma}{}^2 v^\gamma$, $F_{\alpha\beta}F^{1\,\beta\gamma}F_{\gamma\delta}{}^2\bar{F}^{\delta\alpha}$, $v_\alpha F^{\alpha\beta}F_{\beta\gamma}{}^1F^{2\,\gamma\delta}\bar{F}_{\delta\epsilon}v^\epsilon$, or $v_\alpha F^{\alpha\beta}F_{\beta\gamma}{}^1\bar{F}^{\gamma\delta}F_{\delta\epsilon}{}^2v^\epsilon$, or factors obtained from these by permuting the photon field strengths F, F^1, and F^2. A simple direct calculation shows that for free photons propagating along the same direction, the five factors always vanish, irrespective of the orientations of the photon polarizations. We conclude, then, that the term with $2n$ factors k vanishes, so that at most $2n-2$ factors k can be present (for $n \geq 1$) in the term in the photon-splitting matrix element involving $2n+1$ external field factors $\bar{F}$.

Let us now apply this result to the two smallest ring diagrams: the box diagram ($n=0$) and the hexagon diagram ($n=1$). We immediately learn that the box contribution to photon splitting vanishes identically,[6,7] so that the leading diagram which contributes is the hexagon. Further-

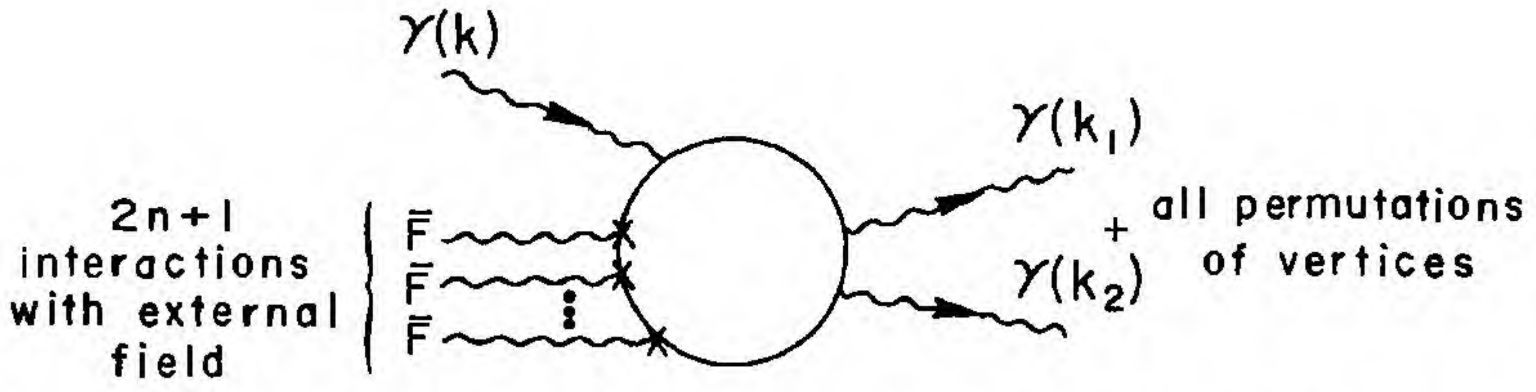


FIG. 1. Ring diagram for photon splitting involving $2n+1$ interactions with the external field.

VOLUME 25, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1970

more, the hexagon contains at most $2\times1-2=0$ factors of k in addition to those contained in the field strengths, which means that the hexagon diagram is given exactly by its constant-field-strength limit, which can in turn be easily calculated from the Heisenberg-Euler[6] effective Lagrangian. Larger ring diagrams will, of course, also be present, but for purposes of rough order-of-magnitude estimates the leading dependence on $\bar{B}/B_{cr}$ and ω/m given by the hexagon should be sufficient. Carrying out the effective-Lagrangian calculation for the hexagon gives the following formulas for the photon-spliting absorption coefficients in the various photon polarization states:

$$\kappa[(\|)\to(\|)_1+(\|)_2]=\frac{\alpha^3}{60\pi^2}\left(\frac{48}{315}\right)^2\left(\frac{\omega}{m}\right)^5\left(\frac{\bar{B}\sin\theta}{B_{cr}}\right)^6 m=0.39\left(\frac{\omega}{m}\right)^5\left(\frac{\bar{B}\sin\theta}{B_{cr}}\right)^6\ \mathrm{cm}^{-1},$$
$$\kappa[(\|)\to(\perp)_1+(\perp)_2]=\frac{\alpha^3}{60\pi^2}\left(\frac{26}{315}\right)^2\left(\frac{\omega}{m}\right)^5\left(\frac{\bar{B}\sin\theta}{B_{cr}}\right)^6 m=0.12\left(\frac{\omega}{m}\right)^5\left(\frac{\bar{B}\sin\theta}{B_{cr}}\right)^6\ \mathrm{cm}^{-1},$$
$$\kappa[(\perp)\to(\|)_1+(\perp)_2]+\kappa[(\perp)\to(\perp)_1+(\|)_2]=2\kappa[(\|)\to(\perp)_1+(\perp)_1]. \tag{7}$$

Here $\alpha=e^2\approx1/137$ is the fine structure constant, θ is the angle between the photon propagation direction $\hat{k}$ and the direction $\hat{b}$ of the external magnetic field, and the linear polarization eigenmodes are labeled $\|$ or $\perp$ according to whether the $\vec{B}$ vector of the eigenmode lies in, or is normal to, the $\hat{k}$-$\hat{b}$ plane. Only formulas for processes involving an even number of $\perp$ photons have been given; the absorption coefficients for processes involving an odd number of $\perp$ photons vanish by a simple CP argument. To see this, we note that for each $\|$ photon, the matrix element will contain a CP-even factor $\vec{B}^{\mathrm{photon}}\cdot\hat{b}$ [the only other possible scalar product, $\vec{B}^{\mathrm{photon}}\cdot\vec{k}$, vanishes by transversality], while for each $\perp$ photon it will contain a CP-odd factor $\vec{E}^{\mathrm{photon}}\cdot\hat{b}$. Since the only scalar not involving the photon fields is $\vec{k}\cdot\hat{b}$, which is CP even, the matrix elements for the odd-$\perp$-photon processes are CP odd, and hence vanish.

So far we have assumed that the photons satisfy the vacuum dispersion relation of Eq. (2). Actually, because of the absorptive processes taking place in the external field, there will be dispersive effects which modify Eq. (2). A simple CP argument shows that the photon eigenmodes remain linearly polarized, with the parallel and perpendicular characters described above, but with the ratio of wave number to frequency changed from unity to

$$k/\omega=n_{\|,\perp}. \tag{8}$$

The indices of refraction $n_{\|,\perp}$ can be calculated from the total absorption coefficients $\kappa_{\|,\perp}$ by Kramers-Kronig (dispersion) relations, with the dominant contribution coming from photopair production [the contribution from photon splitting is smaller by a factor $\sim(\alpha/\pi)^2(\bar{B}\sin\theta/B_{cr})^4$, and can be neglected]. For small $\bar{B}/B_{cr}$ the calculation has been carried out numerically by Toll,[2] who gives curves for $n_{\|,\perp}$ as a function of frequency ω. When we have $\omega<2m$, so that the parameter $x=(\omega/2m)(\bar{B}\sin\theta/B_{cr})$ is also small, Toll's results can be approximated analytically by

$$n_{\|,\perp}=1+\frac{\alpha}{\pi}\left(\frac{\bar{B}\sin\theta}{2B_{cr}}\right)^2 N_{\|,\perp}(x),$$
$$N_{\|}(x)\sim0.18+0.24x^2,$$
$$N_{\perp}(x)\sim0.31+0.44x^2. \tag{9}$$

When dispersive effects are taken into account, the equations of conservation of four-momentum become

$$\omega=\omega_1+\omega_2$$
$$n(\omega)\omega\hat{k}=n(\omega_1)\omega_1\hat{k}_1+n(\omega_2)\omega_2\hat{k}_2, \tag{10}$$

with each n the refractive index appropriate to the respective photon polarization state. These conditions can be simultaneously satisfied only if

$$0\leq\Delta=n(\omega_1)\omega_1+n(\omega_2)\omega_2-(\omega_1+\omega_2)n(\omega_1+\omega_2), \tag{11}$$

in which case the photon propagation directions are not precisely parallel, but rather diverge from one another by small angles $\sim(\Delta/\omega)^{1/2}$. As a result of this nonparallelism, the box diagram is no longer precisely zero, but a careful estimate shows that it is still much smaller than the hexagon. When Δ is negative, the photon-splitting reaction is forbidden. Substituting the indices of refraction of Eq. (9) into Eq. (11) shows, indeed, that for small x the only reaction in Eq. (7) which is kinematically allowed is $(\|)\to(\perp)_1+(\perp)_2$, and that this reaction occurs without restriction on the photon frequencies ω_1 and ω_2.

VOLUME 25, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1970

Table I. Selection rules for photon splitting. *CP*-forbidden reactions are suppressed by a factor $\sim(\alpha/\pi)^2(\bar{B}\sin\theta/B_{\rm cr})^4$ relative to *CP*-allowed cases.

Reaction	*CP* selection rule	Small-x kinematic selection rule
$(\Vert)\to(\Vert)_1+(\Vert)_2$	Allowed	Forbidden
$(\Vert)\to(\Vert)_1+(\perp)_2,\ (\perp)_1+(\Vert)_2$	Forbidden	Allowed
$(\Vert)\to(\perp)_1+(\perp)_2$	Allowed	Allowed
$(\perp)\to(\Vert)_1+(\Vert)_2$	Forbidden	Forbidden
$(\perp)\to(\Vert)_1+(\perp)_2,\ (\perp)_1+(\Vert)_2$	Allowed	Forbidden
$(\perp)\to(\perp)_1+(\perp)_2$	Forbidden	Forbidden

The various polarization selection rules for photon splitting are summarized in Table I. Because of the small nonparallelism of the photons in the kinematically allowed regions, the "*CP*-forbidden" reactions are not precisely forbidden, but are down by a factor $\sim(\alpha/\pi)^2(\bar{B}\sin\theta/B_{\rm cr})^4$ relative to the "*CP*-allowed" cases. We see that for small x, all reactions by which perpendicularly polarized photons might split are kinematically forbidden, while parallel-polarized photons split predominantly into perpendicularly polarized photons. Hence photon splitting provides a mechanism for the production of linearly polarized γ rays.[9]

To conclude, let us briefly discuss the possible application of our results to pulsars. We assume that the hexagon-diagram absorption coefficients in Eq. (7) can be used for order-of-magnitude estimates even when the parameters $\bar{B}/B_{\rm cr}$ and ω/m are of order unity.[10] Taking, for illustration, $\bar{B}/B_{\rm cr}\sim\omega/m\sim\sin\theta\sim1$, we find $\kappa[(\Vert)\to(\perp)_1+(\perp)_2]\sim0.1\ \text{cm}^{-1}$. This gives 10^5 absorption lengths in the characteristic distance $R_{\rm pulsar}\sim10^6$ cm over which the trapped magnetic field has its maximum strength, indicating that photon splitting can be an important absorption mechanism for γ rays emitted near the pulsar surface. Before we can apply the kinematic polarization selection rules to the pulsar problem, two questions must be dealt with. First, since Toll's curves for the indices of refraction were obtained assuming $\bar{B}/B_{\rm cr}$ small, an extrapolation is involved in extending the selection rules forbidding perpendicular-photon decay and parallel-photon decay into parallel photons to the region where $\bar{B}/B_{\rm cr}$ is of order unity. However, Toll's photopair production curves show that $\kappa_\perp>\kappa_\Vert$ when $\bar{B}/B_{\rm cr}$ is unity. By combining this fact with the Kramers-Kronig relations, one easily sees that the selection rules in Table I hold as long as $\omega<2m$, and hence the extrapolation is justified. Second, one expects a plasma to be present near the pulsar surface which will contribute additional dispersive terms to the inequality of Eq. (11). For a plasma-electron density of 10^{17}-10^{-19} cm^{-3} (in rough accord with current pulsar models[10]), a detailed estimate shows that plasma-induced splitting of perpendicularly polarized photons occurs with an absorption coefficient of at most 10^{-9}-10^{-7} times the absorption coefficient for the allowed reaction $(\Vert)\to(\perp)_1+(\perp)_2$, and therefore will be completely negligible.[11] We conclude that if γ rays in the range 0.5-1 MeV are emitted near the pulsar surface, and if the surface magnetic field is as large as $B_{\rm cr}$, only those gammas with perpendicular polarization will escape. A distant observer would see linearly polarized gammas, with their $\vec{B}$ vector perpendicular to the plane containing the line of sight and the traversed pulsar magnetic field.[12]

A detailed account of the calculations summarized here will be presented elsewhere.[10] The authors wish to thank P. Goldreich (who first brought this problem to our attention), E. P. Lee, and M. Rassbach for informative conversations. After this manuscript was completed, we learned that some of our results have been obtained independently by Z. Bialynicka-Birula and I. Bialynicki-Birula.[13]

---

*Present address: California Institute of Technology, Pasadena, Calif.

[1]For a recent view, see F. Pacini, "Nentron Stars, Pulsar Radiation and Supernova Remnants," to be published. We use unrationalized Gaussian units, with $\hbar=c=1$.

[2]J. Toll, dissertation, Princeton University, 1952 (unpublished).

[3]The phase space for a photon to split into three or more photons vanishes.

[4]In a pulsar, the field $\bar{B}$ varies over a characteristic distance of $R_{\rm pulsar}\sim10^6$ cm, but this can be shown to have a negligible effect on our results.

[5]We need not consider contractions involving the

VOLUME 25, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1970

antisymmetric tensor $\epsilon_{\alpha\beta\gamma\delta}$, because by parity the number of such factors must be even, and they can be eliminated pairwise in terms of Kronecker deltas by means of the identity

$$\epsilon_{\alpha\beta\gamma\delta}\epsilon_{\alpha'\beta'\gamma'\delta'}=\sum_{\mathrm{Perm}(\alpha'\beta'\gamma'\delta')}(-1)^P g_{\alpha\alpha'}g_{\beta\beta'}g_{\gamma\gamma'}g_{\delta\delta'}.$$

[6]This result for the box diagram has been obtained independently by M. Rassbach.

[7]This disagrees with the conclusion of V. G. Skobov, Zh. Eksp. Teor. Fiz. 35, 1315 (1958) [Sov. Phys. JETP 8, 919 (1959)], who failed to make a properly gauge-invariant calculation of the box. Skobov's result is quoted in the review article of T. Erber, Rev. Mod. Phys. 38, 626 (1966).

[8]W. Heisenberg and H. Euler, Z. Phys. 38, 714 (1936).

[9]Toll (Ref. 2) points out that in a frequency interval $\Delta\omega \sim e\bar{B}/m$ just above the photopair threshold at $\omega = 2m$, the photopair process acts as a linear polarizer of the opposite sense, absorbing photons of perpendicular polarization, but not those of parallel polarization.

[10]For further discussion, see S. L. Adler, to be published.

[11]In a plasma, the propagation eigenmodes become elliptically polarized, but are still "almost plane ∥" and "almost plane ⊥" in nature. Faraday rotation, which arises from interference between two unattenuated propagation eigenmodes of different phase velocities, cannot occur in our case since only the "almost plane ⊥" eigenmode propagates without attenuation. As a result of photon splitting the "almost plane ∥" eigenmode is rapidly absorbed.

[12]Solid-state linear-polarization analyzers for gammas in this energy range have been described by G. T. Ewan *et al.*, Phys. Lett. 29B, 352 (1969).

[13]Z. Bialynicka-Birula and I. Bialynicki-Birula, to be published.

ANNALS OF PHYSICS: 67, 599–647 (1971)

Photon Splitting and Photon Dispersion in a Strong Magnetic Field

STEPHEN L. ADLER

Institute for Advanced Study, Princeton, New Jersey 08540

Received January 27, 1971

We determine the refractive indices for photon propagation, and the absorption coefficient and polarization selection rules for photon splitting, in a strong constant magnetic field. Results are presented both in the effective Lagrangian (low frequency) approximation and in a more accurate approximation which exactly sums the vacuum polarization ring diagrams, neglecting only internal virtual photon radiative corrections. Our principal conclusion is that photon splitting can provide a mechanism for the production of linearly polarized gamma rays.

1. INTRODUCTION AND SUMMARY

Recent work on pulsars has suggested the presence of trapped magnetic fields within an order of magnitude (in either direction) of the electrodynamic critical field $B_{CR} = m^2/e = 4.41 \cdot 10^{13}$ gauss[1] (with m and e, respectively, the electronic mass and charge)[1]. In such intense fields, electrodynamic processes which are unobservable in the laboratory can become important. One such process, for photons with energy $\omega > 2m$, is photo-pair production, for which both the photon absorption coefficient (inverse absorption length) and the corresponding vacuum dispersion have been calculated by Toll [2]. For $\omega < 2m$, the photo-pair process is kinematically forbidden, and the only photon absorption mechanism in the absence of matter is photon splitting, i.e.,

$$\gamma(k) + \text{external magnetic field} \to \gamma(k_1) + \gamma(k_2). \tag{1}$$

We give in this paper detailed calculations of the absorption coefficient and polarization selection rules for this reaction, in the case of a constant and spatially uniform external magnetic field $\bar{B}$. A summary of our results, and a brief discussion of their possible application to pulsars, have already been given in [3].

In Section 2, we consider photon splitting in the absence of dispersion, so that the three photons are strictly collinear. We find that the box diagram matrix

[1] We use unrationalized Gaussian units, with $\hbar = c = 1$.

element vanishes, while the hexagon matrix element is nonvanishing and is given exactly by its constant-field-strength (small ω/m) limit. Using the Heisenberg–Euler effective Lagrangian [4], we calculate the sum of the constant field strength limits of all ring diagrams for photon splitting involving arbitrary numbers of interactions with the external field. Then, we use proper-time techniques to exactly calculate the photon splitting matrix element to all orders in the external field, *without* the restriction to constant photon field strengths. (The details of this latter calculation, which neglects only internal virtual photon radiative corrections, are given in Appendix 1). The magnitude of the resulting photon splitting absorption coefficient, for photon propagation normal to the external field, is $\kappa \sim 0.1(\bar{B}/B_{CR})^6 (\omega/m)^5 \text{ cm}^{-1}$. Thus, if pulsar fields are as large as B_{CR}, for photon frequencies ω of order m, there are many photon splitting absorption lengths in a characteristic pulsar distance of 10^6 cm. A comparison of the photon splitting and photo-pair absorption coefficients shows that when the photo-pair process is kinematically allowed it dominates over photon splitting as a photon absorption mechanism. To complete our discussion of the no-dispersion case, we estimate the leading corrections arising from the fact that the magnetic field $\bar{B}$ is not strictly uniform, but varies over characteristic distances of order 10^6 cm. A nonuniform magnetic field can transfer momentum to the photons, with the result that the final photons emerge at small but finite angles with respect to the initial photon direction. This means that the argument for the vanishing of the box diagram no longer holds, but an estimate shows that the resulting box diagram contribution is negligibly small compared with the hexagon diagram absorption coefficient.

In Section 3 we discuss dispersion effects and polarization selection rules. Because of photon absorption processes which take place in the presence of the external magnetic field $\bar{B}$, the vacuum in the presence of the field $\bar{B}$ acquires an index of refraction n, and the photon dispersion relation is modified from $k/\omega = 1$ to $k/\omega = n$. There are actually two different indices of refraction n corresponding to the two photon propagation eigenmodes. A general argument based on the CP invariance of electrodynamics shows that the eigenmodes are linearly polarized, with the **B**-vector of the eigenmode either parallel to ($\|$ mode) or perpendicular to ($\perp$ mode) the plane containing the external field and the direction of propagation. The indices of refraction $n_{\|,\perp}$ can be calculated in the constant field strength (small ω/m) limit from the Heisenberg–Euler effective lagrangian, and can be calculated without the constant field strength restriction, by the proper time methods of Appendix 1. They can also be obtained from the absorption coefficients $\kappa_{\|,\perp}$ by Kramers–Kronig relations, with the dominant contribution coming from photo-pair production. When dispersive effects are taken into account, we find that energy-momentum conservation in the photon splitting process can be satisfied only if the inequality $0 \leqslant \Delta = n(\omega_1)\,\omega_1 + n(\omega_2)\,\omega_2 - n(\omega_1 + \omega_2)(\omega_1 + \omega_2)$, holds, with $\omega = \omega_1 + \omega_2$ and with each n the refractive index appropriate to the

respective photon polarization state. The photon polarization directions are no longer precisely parallel, but rather diverge from one another by small angles of order $(\Delta/\omega)^{1/2}$. Again, as a result of this nonparallelism, the box diagram is no longer precisely zero, but a careful estimate shows that it is only an order $\alpha(\alpha = e^2 =$ fine structure constant) correction to the hexagon. When $\Delta < 0$, the photon splitting reaction is forbidden. Using our expressions for the refractive indices, we analyze the sign of Δ for the various photon polarization cases. We find that when ω is below the pair production threshold at $2m$, only the photon splitting reactions $(\|) \to (\perp)_1 + (\perp)_2$ and $(\|) \to (\|)_1 + (\perp)_2, (\perp)_1 + (\|)_2$ are kinematically allowed. Furthermore, when the small angles between the photon propagation directions are neglected, a simple *CP*-invariance argument shows that the photon splitting reactions involving an odd number of $(\perp)$ photons are forbidden. Hence the only allowed polarization case is $(\|) \to (\perp)_1 + (\perp)_2$, indicating that photon splitting provides a mechanism for the production of polarized photons: perpendicular-polarized photons do not split, and parallel-polarized photons split predominantly into perpendicular photons.

Finally, in Section 4 we discuss corrections to our results arising when a plasma with electron density n_e is present in the region containing the strong magnetic field. We show that for n_e of order $10^{17} - 10^{19}$ cm^{-3} (in rough accord with current pulsar models) the plasma-induced splitting of perpendicular-polarized photons occurs with an absorption coefficient of at most $10^{-9} - 10^{-7}$ times the absorption coefficient for the allowed reaction $(\|) \to (\perp)_1 + (\perp)_2$, and hence will be completely negligible.

2. No-Dispersion Case[2]

A. *Kinematics*

We consider photon splitting in the presence of a time-independent, spatially uniform external magnetic field $\vec{B}$. Because of the constancy of the external field it can absorb no four-momentum, and as a result the four-vectors k, k_1, k_2 of the initial and final photons must satisfy energy and momentum conservation by themselves,

$$k = \omega(1, \hat{k}) = k_1 + k_2 = \omega_1(1, \hat{k}_1) + \omega_2(1, \hat{k}_2). \tag{2}$$

It is easily seen that this condition can be satisfied only if the propagation directions of the three photons are identical,

$$\hat{k} = \hat{k}_1 = \hat{k}_2, \tag{3}$$

[2] We follow the metric and other notational conventions of J. D. Bjorken and S. D. Drell, "Relativistic Quantum Fields," McGraw-Hill, New York, 1965, pp. 377–390.

external field direction. The vanishing of all matrix elements involving an odd number of perpendicular photons results from the CP invariance of quantum electrodynamics, as will be explained in detail below.

Finally, doing the phase space integrals gives us the following expressions for the photon splitting absorption coefficients:

$$\begin{aligned}
&\kappa[(\|) \to (\|)_1 + (\|)_2] \\
&\quad = \frac{1}{32\pi\omega^2} \int_0^\omega d\omega_1 \int_0^\omega d\omega_2\, \delta(\omega - \omega_1 - \omega_2)\, |\mathscr{M}[(\|) \to (\|)_1 + (\|)_2]|^2 \\
&\quad = \frac{\alpha^6}{2\pi^2} \frac{\bar{B}^6 \sin^6\theta}{m^{16}}\, C_1(\bar{B}/B_{CR})^2\, J, \\
&\kappa[(\|) \to (\perp)_1 + (\perp)_2] \\
&\quad = \frac{1}{32\pi\omega^2} \int_0^\omega d\omega_1 \int_0^\omega d\omega_2\, \delta(\omega - \omega_1 - \omega_2)\, |\mathscr{M}[(\|) \to (\perp)_1 + (\perp)_2]|^2 \\
&\quad = \frac{\alpha^6}{2\pi^2} \frac{\bar{B}^6 \sin^6\theta}{m^{16}}\, C_2(\bar{B}/B_{CR})^2\, J, \\
&\kappa[(\perp) \to (\|)_1 + (\perp)_2] + \kappa[(\perp) \to (\perp)_1 + (\|)_2] \\
&\quad = \frac{1}{32\pi\omega^2} \int_0^\omega d\omega_1 \int_0^\omega d\omega_2\, \delta(\omega - \omega_1 - \omega_2) \\
&\qquad \times \{|\mathscr{M}[(\perp) \to (\|)_1 + (\perp)_2]|^2 + |\mathscr{M}[(\perp) \to (\perp)_1 + (\|)_2]|^2\} \\
&\quad = \frac{\alpha^6}{2\pi^2} \frac{\bar{B}^6 \sin^6\theta}{m^{16}}\, 2C_2(\bar{B}/B_{CR})^2\, J,
\end{aligned} \tag{23}$$

with

$$J = \int_0^\omega d\omega_1 \int_0^\omega d\omega_2\, \delta(\omega - \omega_1 - \omega_2)\, \omega_1^2\omega_2^2 = \frac{\omega^5}{30}. \tag{24}$$

Eqs. (15), (17), (23) and (24) constitute our results for photon splitting in the small ω/m limit when dispersive effects are neglected [3, 8]. We will see below that when dispersive effects are taken into account, the reactions $(\|) \to (\|)_1 + (\|)_2$, $(\perp) \to (\|)_1 + (\perp)_2$ and $(\perp) \to (\perp)_1 + (\|)_2$ are kinematically forbidden, while the reaction $(\|) \to (\perp)_1 + (\perp)_2$ still occurs with the absorption coefficient given by Eq. (23).

D. *Exact Calculation*

As we have noted, the effective Lagrangian formulas of Eq. (23) give the sum of ring diagrams shown in Fig. 4, in the limit of constant photon field strength

(small ω/m). It is possible, by using the proper-time techniques developed by Schwinger [7], to exactly calculate this sum of ring diagrams, without the small ω/m restriction. (Virtual photon radiative corrections to the ring diagrams, such as shown in Fig. 6, are still neglected, but these are expected to be strictly an order

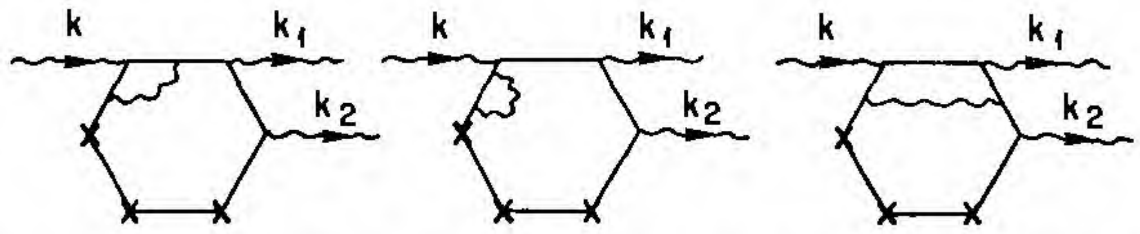


FIG. 6. Typical virtual photon radiative corrections to the hexagon diagram for photon splitting (cf. Fig. 3).

α correction to our results.) Leaving calculational details to Appendix 1, we present here only the results for the kinematically allowed reaction $(\|) \to (\perp)_1 + (\perp)_2$. We find that the matrix element $\mathscr{M}[(\|) \to (\perp)_1 + (\perp)_2]$ appearing in Eq. (23) is now given by the rather complicated expression

$$\mathscr{M}[(\|) \to (\perp)_1 + (\perp)_2] = \frac{\alpha^3}{2\pi^2 m^8} [\bar{B} \sin\theta (4\pi)^{1/2}]^3\, \omega\omega_1\omega_2 C_2[\omega \sin\theta, \omega_1 \sin\theta, \omega_2 \sin\theta, \bar{B}], \tag{25a}$$

$$C_2[\omega, \omega_1, \omega_2, \bar{B}] = \frac{m^8}{16\omega\omega_1\omega_2} \int_0^\infty \frac{s\,ds \exp(-m^2 s)}{(es\bar{B})^2 \sinh(es\bar{B})} \Big\{-8 \int_0^s dt \int_0^t du \times [\omega\omega_1\omega_2 \tilde{A} + (\omega_1 - \omega_2)\,\omega_1\omega_2 \tilde{B} + s^{-1}\omega\tilde{C} + s^{-1}(\omega_1 - \omega_2)\,\tilde{D}] + 8 \int_0^s dt\, \omega \tilde{E}\Big\},$$

$$\tilde{A} = \{\exp[\omega_2^2 R(s,t) + \omega_1^2 R(s,u)] + \exp[\omega_1^2 R(s,t) + \omega_2^2 R(s,u)]\} \times \exp[\omega_1\omega_2\, \Delta(s,t,u)] \sinh[e\bar{B}(t-u)] \times \Big\{\frac{C_+(s,t,u) - \cosh(e\bar{B}s)}{\sinh(e\bar{B}s)} \sinh[e\bar{B}(s-t+u)] + 2 \sinh[e\bar{B}(s-t)] \sinh(e\bar{B}u)\Big\},$$

$$\tilde{B} = \{\exp[\omega_2^2 R(s,t) + \omega_1^2 R(s,u)] - \exp[\omega_1^2 R(s,t) + \omega_2^2 R(s,u)]\} \times \exp[\omega_1\omega_2\, \Delta(s,t,u)] \sinh[e\bar{B}(t-u)] \frac{C_-(s,t,u)}{\sinh(e\bar{B}s)} \sinh[e\bar{B}(s-t+u)],$$

$$\tilde{C} = \{[\exp[\omega_2^2 R(s,t) + \omega_1^2 R(s,u)] + \exp[\omega_1^2 R(s,t) + \omega_2^2 R(s,u)]] \times \exp[\omega_1\omega_2\, \Delta(s,t,u)] - 2\} \Big[\sinh(e\bar{B}s) + \frac{C_+(s,t,u) - \cosh(e\bar{B}s)}{\sinh(e\bar{B}s)} \cosh(e\bar{B}s)\Big],$$

$$\tilde{D} = \{\exp[\omega_2{}^2 R(s, t) + \omega_1{}^2 R(s, u)] - \exp[\omega_1{}^2 R(s, t) + \omega_2{}^2 R(s, u)]\}$$
$$\times \exp[\omega_1 \omega_2 \Delta(s, t, u)]\, C_-(s, t, u) \coth(e\bar{B}s),$$

$$\tilde{E} = \{\exp[\omega^2 R(s, t)] - 1\}$$
$$\times \left[\sinh(e\bar{B}s) + \frac{\cosh[(e\bar{B}s)(2t/s - 1)] - \cosh(e\bar{B}s)}{\sinh(e\bar{B}s)} \cosh(e\bar{B}s)\right],$$

with

$$C_\pm(s, t, u) = \frac{1}{2}\left\{\cosh\left[(e\bar{B}s)\left(\frac{2u}{s} - 1\right)\right] \pm \cosh\left[(e\bar{B}s)\left(\frac{2t}{s} - 1\right)\right]\right\},$$

$$R(s, t) = \frac{1}{2}\left[2t\left(1 - \frac{t}{s}\right) + \frac{\cosh[e\bar{B}s(2t/s - 1)] - \cosh(e\bar{B}s)}{e\bar{B}\sinh(e\bar{B}s)}\right],$$

$$\Delta(s, t, u) = 2u\left(1 - \frac{t}{s}\right) - \frac{\sinh(2e\bar{B}u)}{2e\bar{B}}\left[1 - \frac{\sinh[(e\bar{B}s)(2t/s - 1)]}{\sinh(e\bar{B}s)}\right]$$
$$+ \frac{[1 - \cosh(2e\bar{B}u)]}{2e\bar{B}}\left[\frac{\cosh[(e\bar{B}s)(2t/s - 1)] - \cosh(e\bar{B}s)}{\sinh(e\bar{B}s)}\right]. \tag{25b}$$

We make some remarks on various features of this formula:

(1) Bose symmetry, which states that $\mathscr{M}[(\|) \to (\perp)_1 + (\perp)_2]$ is symmetric under the interchange $\omega_1 \leftrightarrow \omega_2$, is explicitly evident.

(2) The small ω/m limit of Eq. (25) is obtained by replacing $\tilde{A}$ by the zeroth order term and $\tilde{C}$, $\tilde{D}$ and $\tilde{E}$ by the (leading) second order terms in their respective expansions in powers of photon frequency. (The term $\tilde{B}$ does not contribute to the constant field-strength limit.) Doing the u and t integrations then shows that, in the small ω/m limit, $C_2[\omega, \omega_1, \omega_2, \bar{B}]$ reduces to the effective Lagrangian expression $C_2(\bar{B}/B_{CR})$, in just the form given in Eq. (15).

(3) From Eq. (25b), we easily see that $R(s, t)$ and $\Delta(s, t, u)$ are even functions of the external field $\bar{B}$, both of which vanish like $\bar{B}^2$ when $\bar{B}$ is small. Therefore, the small $\bar{B}$-limit of Eq. (25) is obtained by expanding out the curly-bracketed exponential terms in $\tilde{A}, \ldots, \tilde{E}$, just as in the small ω/m case. The leading small $\bar{B}$ contribution to $C_2[\omega, \omega_1, \omega_2, \bar{B}]$ is of order $\bar{B}^3$, as expected for the hexagon diagram, and is found to be frequency-independent, confirming our above-stated result that the hexagon diagram is given exactly by its constant field strength limit. When higher order terms in the expansions of the curly-bracketed exponentials are kept, we see that the general term in C_2 with $2n$ powers of photon frequency contains *at least* $2n + 3$ powers of $\bar{B}$, in agreement with the theorem of Subsection 2*B*.

(4) Because Eq. (25) contains infinite multiple integrals, it is necessary to look carefully at the question of convergence. By making the rescalings

$$\begin{aligned} t &= \frac{s}{2}(1+v), & \int_0^s dt &= \frac{s}{2}\int_{-1}^{1} dv, \\ u &= \frac{s}{2}(1+z), & \int_0^t du &= \frac{s}{2}\int_{-1}^{v} dz, \end{aligned} \tag{26}$$

the t and u integrals are transformed into integrals over finite intervals, with only the s integral still extending to infinity. To examine the large-s behavior of the rescaled integrand, we use the fact that when $s \to \infty$ with t/s and u/s fixed, the quantities $R(s, t)$ and $\Delta(s, t, u)$ are simply approximated by

$$\begin{aligned} R(s, t) &\approx t(1 - t/s) + \text{finite}, \\ \Delta(s, t, u) &\approx 2u(1 - t/s) + \text{finite}. \end{aligned} \tag{27}$$

On substituting these expressions into Eq. (25), and making similar large-s approximations in the factors multiplying the curly brackets, we obtain the following result for the region of convergence: When ω is below the pair production threshold at $\omega = 2m$, Eq. (25) converges at least as fast as

$$\int^{\infty} ds\, e^{-s(m^2-\omega^2/4)} \times (\text{power of } s) \tag{28}$$

for all values of the secondary photon frequencies ω_1 and ω_2. When ω is above the pair production threshold, there are values of ω_1 and ω_2 for which the integral diverges. Thus, Eq. (25) gives a valid expression for the photon splitting matrix element only in the photon energy region where this matrix element is real, but fails when the photon splitting matrix element becomes complex, as a result of absorptive contributions such as the one pictured in Fig. 7. This failure is no real problem, since we will see below that when $\omega > 2m$, the photopair production process is a far more important photon absorption mechanism than is photon

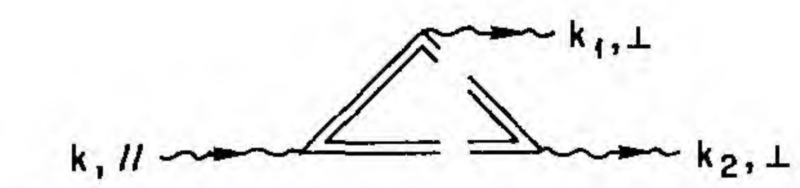


FIG. 7. Lowest-lying absorptive contribution to photon splitting, corresponding to an electron-positron intermediate state. The doubled lines indicate the presence of interactions to all orders with the external field $\bar{F}$. (Such interactions were individually denoted by an $\times$ in Figs. 2, 3, 4, and 6.)

splitting, and so an exact formula for the photon splitting matrix element in this region is not really of interest.

(5) As we noted above, in the limit of small $\bar{B}$ the matrix element $C_2[\omega, \omega_1, \omega_2, \bar{B}]$ vanishes as $\bar{B}^8$. However, a glance at Eq. (25) shows that there are individual terms which vanish with lower powers of $\bar{B}$; the $\bar{B}^3$ behavior is a result of cancellations. It is easy to see, though, that these cancellations are entirely contained in the integrands $\tilde{A}, \ldots, \tilde{E}$, and do not involve the s, t, u integrations. As a result, reliable numerical results for the photon splitting matrix element and absorption coefficient can be obtained with rather coarse integration meshes. Typical numerical results are shown in Fig. 8, which gives the ratio of the exact photon

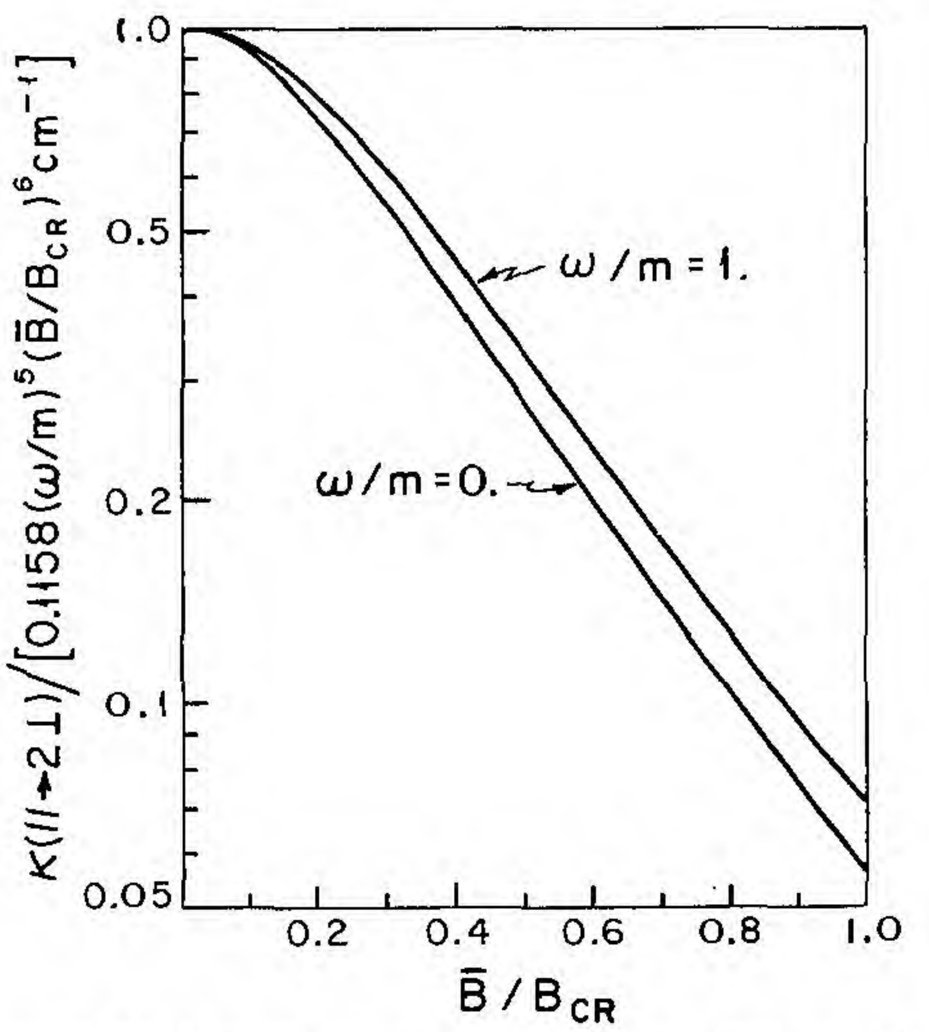


FIG. 8. Ratio of the exact photon splitting absorption coefficient $\kappa[(\parallel) \to (\perp)_1 + (\perp)_2]$ to the hexagon diagram value for the same quantity.

splitting absorption coefficient calculated from Eq. (25) to the absorption coefficient obtained from the hexagon diagram alone (Eq. (23) with $C_2(\bar{B}/B_{CR})$ replaced by $C_2(0) = 6 \cdot 13/945$). The upper curve is the result for $\omega/m = 1$, while the lower curve, giving the low frequency limit, is identical to a plot of $C_2(\bar{B}/B_{CR})^2$ (cf. Fig. 5 which gives a plot of $C_2(\bar{B}/B_{CR})$). We see that in the range $0 \leqslant \bar{B}/B_{CR} \leqslant 1$, $0 \leqslant \omega/m \leqslant 1$ the leading power dependence given by the hexagon diagram alone suffices for rough order of magnitude estimates. Furthermore, the effective Lagrangian calculation of $C_2(\bar{B}/B_{CR})$ gives the bulk of the corrections coming from higher ring diagrams, with the ω-dependent effects contained in the complicated formulas of Eq. (25) (i.e., the spread between the two curves in Fig. 8) being fairly small.

As expected, substituting the Taylor expansions

$$\begin{aligned} K^{\parallel}(z) &= -8z^2/45 + O(z^4), \\ K^{\perp}(z) &= -14z^2/45 + O(z^4), \end{aligned} \tag{50}$$

into Eq. (49) gives back the box diagram result of Eq. (46b) in the limit of small $\bar{B}/B_{CR}$. The calculation can be further improved by using the proper-time techniques discussed in Appendix 1 to exactly sum the ring diagrams involving arbitrary numbers of interactions with the external field $\bar{B}$, without the restriction to small ω/m. This gives the following formulas for the refractive indices,

$$\begin{aligned} n^{\parallel,\perp} &= 1 - \frac{1}{2}\sin^2\theta A^{\parallel,\perp}[\omega \sin\theta, \bar{B}], \\ A^{\parallel,\perp}[\omega, \bar{B}] &= \frac{\alpha}{2\pi}\int_0^\infty \frac{ds}{s^2}\exp(-m^2 s)\int_0^s dt\, \exp[\omega^2 R(s,t)]\, J^{\parallel,\perp}(s,v), \\ v &= 2t/s - 1, \\ J^{\parallel}(s,v) &= \frac{-e\bar{B}s\cosh(e\bar{B}sv)}{\sinh(e\bar{B}s)} + \frac{e\bar{B}sv\sinh(e\bar{B}sv)\coth(e\bar{B}s)}{\sinh(e\bar{B}s)} \\ &\quad - \frac{2e\bar{B}s[\cosh(e\bar{B}sv) - \cosh(e\bar{B}s)]}{\sinh^3(e\bar{B}s)}, \\ J^{\perp}(s,v) &= \frac{e\bar{B}s\cosh(e\bar{B}sv)}{\sinh(e\bar{B}s)} - e\bar{B}s\coth(e\bar{B}s)\left[1 - v^2 + v\frac{\sinh(e\bar{B}sv)}{\sinh(e\bar{B}s)}\right], \end{aligned} \tag{51}$$

and with $R(s, t)$ given by Eq. (25b). When ω is set equal to zero, the integral over t (or equivalently, over v) is readily done, giving $\int_0^1 dv\, J^{\parallel,\perp}(s, v) = K^{\parallel,\perp}(e\bar{B}s)$, and so Eq. (51) reduces directly to Eq. (49). To examine the region of convergence of Eq. (51), we substitute Eq. (27) for $R(s, t)$ and replace $J^{\parallel,\perp}(s, v)$ by their dominant large-s behavior, giving

$$n^{\parallel\perp} - 1 \propto \int^\infty ds \int_0^s dt\, \exp[W^{\parallel,\perp}] \times (\text{power of } s), \tag{52}$$

with

$$\begin{aligned} W^{\parallel} &= \omega^2 t\left(1 - \frac{t}{s}\right) - m^2 s + e\bar{B}(|2t - s| - s), \\ W^{\perp} &= \omega^2 t\left(1 - \frac{t}{s}\right) - m^2 s. \end{aligned} \tag{53}$$

Maximizing $W^{\perp}$ and $W^{\parallel}$ with respect to t, we find that the integral for $n^{\perp}$ con-

verges for $\omega < 2m$, while that for $n^{\|}$ converges in the somewhat larger region $\omega < m[1 + (1 + 2\bar{B}/B_{CR})^{1/2}]$. As Toll [2] has shown, these are just the thresholds for photopair production by perpendicular-polarized and by parallel-polarized photons, respectively. Thus, Eq. (51) is valid when the refractive indices are real, but fails when the refractive indices become complex, as a result of the presence of the absorptive processes pictured in Fig. 12.

FIG. 12. Lowest-lying absorptive contributions to the refractive indices for parallel (‖)- and perpendicular (⊥)-polarized photons. The doubled lines indicate the presence of interactions to all orders in the external field $\bar{F}$.

An alternative method for calculating the indices of refraction uses the fact that for each eigenmode the refractive index n is related to the corresponding absorption coefficient κ by the Kramers-Kronig (dispersion) relation

$$n(\omega) = 1 + \frac{P}{\pi} \int_0^\infty \frac{\kappa(\omega')\, d\omega'}{\omega'^2 - \omega^2}. \tag{54}$$

The dominant contribution to Eq. (54) comes from photo-pair production, which gives

$$(n-1)^{\text{pair}} \sim \frac{\alpha}{\pi}\left(\frac{\bar{B}\sin\theta}{B_{CR}}\right)^2 = \frac{\alpha^2\bar{B}^2\sin^2\theta}{\pi m^4}, \tag{55}$$

as found in Eq. (46b); in comparison to this, the photon-splitting contribution, which can be estimated to be of order

$$(n-1)^{\text{photon splitting}} \sim \left(\frac{\alpha}{\pi}\right)^8 \left(\frac{\bar{B}\sin\theta}{B_{CR}}\right)^2, \tag{56}$$

can be neglected. The contribution of other absorptive processes to the index of refraction will be even smaller. Thus, substituting into Eq. (54) the actual thresholds for photopair production by parallel and perpendicular polarized photons, we get

$$\begin{aligned} n_{\|}(\omega) &= 1 + \frac{P}{\pi} \int_{m[1+(1+2\bar{B}/B_{CR})^{1/2}]}^\infty \frac{\kappa_{\|}^{\text{pair}}(\omega')\, d\omega'}{\omega'^2 - \omega^2}, \\ n_{\perp}(\omega) &= 1 + \frac{P}{\pi} \int_{2m}^\infty d\omega' \frac{\kappa_{\perp}^{\text{pair}}(\omega')\, d\omega'}{\omega'^2 - \omega^2}. \end{aligned} \tag{57}$$

We will not actually evaluate these formulas numerically, but will make use below of the computational results of Toll [2], which show that for $\bar{B}/B_{CR} \lesssim 1$ the perpen-

Region 4. In this region $| n^{\mathrm{TOT}} - 1 |$ can become as big as unity, so we have

$$\frac{\text{forbidden decays}}{\text{allowed decay}} \sim \frac{\omega^2(\Omega_C{}^i)^2\, 10^{-7}}{\omega^5/30} \left(\frac{B_{CR}}{\bar{B}\sin\theta}\right)^4 \sim 10^{-11}. \tag{87c}$$

Region 5. According to Eq. (74), $n^{\mathrm{TOT}} - 1$ can become arbitrarily large in region 5 because ω_1 can come arbitrarily close to $\Omega_C{}^e$. But thermal effects, which have so far been neglected, prevent the effective value of $| \omega_1 - \Omega_C{}^e |$ in Eq. (74) from becoming greater than the thermal spread $\Delta\Omega_C{}^e$ in the electron cyclotron frequency, which is

$$\Delta\Omega_C{}^e \sim \left\langle \frac{e\bar{B}}{m} - \frac{e\bar{B}}{(\mathbf{p}^2 + m^2)^{1/2}} \right\rangle_{AV} \approx \Omega_C{}^e \frac{\langle \mathbf{p}^2/2m\rangle_{AV}}{m} = \Omega_C{}^e \frac{(3/2)\, kT}{m}. \tag{88}$$

Even for T as low as $10^{-1}\,{}^\circ K$ we have $\Delta\Omega_C{}^e \gtrsim 2 \cdot 10^{-12} m$, which limits $| n^{\mathrm{TOT}} - 1 |$ to 10^{-2} at most and makes Eq. (86) smaller than one. Hence we get

$$\frac{\text{forbidden decay}}{\text{allowed decay}} \sim \frac{\omega^2(\Omega_C{}^e)^2\, 10^{-7}}{\omega^5/30} \sim 3 \cdot 10^{-8} < 10^{-7}. \tag{87d}$$

We conclude, then, that the reactions of Eq. (83) have absorption coefficients which are at most 10^{-7} of the absorption coefficient of the allowed reaction $(\|) \to (\perp)_1 + (\perp)_2$. A more detailed analysis shows that this upper band of 10^{-7} applies particularly to the decay $(\|) \to (\|)_1 + (\|)_2$; in the case of all of the decays of an initially perpendicular photon, the upper bound can be reduced at least another two orders of magnitude, to 10^{-9}. Although all of our estimates have assumed $\bar{B}/B_{CR} \approx 0.1$, as $\bar{B}$ is further increased the electrodynamic terms in Eq. (81) rapidly increase in size relative to the plasma terms, causing our upper bound to get even smaller. On the other hand, if the plasma density n_e is increased by a moderate factor the upper bound only increases proportionally. For example a factor of 100 increase in density to $n_e \sim 10^{19}\,\mathrm{cm}^{-3}$ leads to an upper bound on perpendicular photon splitting of 10^{-7} relative to the allowed case. *We conclude, that plasma induced violations of our selection rule against perpendicular photon decay should be negligibly small.*

APPENDIX I: Proper-Time Calculation of the Refractive Indices and Photon Splitting Matrix Element

We outline here the calculations leading to Eq. (51) for the refractive indices and Eq. (25) for the photon splitting matrix element. As noted in the text, we work to all orders in the strong external field $\bar{B}$, without restrictions on ω/m (other than

those needed to assure convergence of the final formulas), but we neglect virtual photon radiative corrections. This means that we are dealing with the problem of vacuum polarization effects produced by a c-number (unquantized) electromagnetic field, which is a superposition of the strong, constant field $\bar{B}$ and of the plane wave fields of the photons. Diagrammatically, we are discussing the problem of a single virtual electron loop, with arbitrary numbers of interactions with the external fields and either with two photon vertices (refractive index calculation, Figs. 9 and 11) or with three photon vertices (photon splitting calculation, Fig. 4).

Very powerful techniques for dealing with the vacuum polarization of c-number fields were developed some time ago by Schwinger [7], and we will follow his methods quite closely. We begin by finding an expression for the electromagnetic current density $\langle j_\mu(x)\rangle$ induced by vacuum polarization effects at point x when an external c-number electromagnetic field A_μ is applied to the vacuum. Denoting the electron field by ψ, and the electron-positron vacuum expectation by $\langle\ \rangle_0$, we have

$$\begin{aligned} j_\mu(x) &= \tfrac{1}{2}e[\bar{\psi}(x), \gamma_\mu\psi(x)], \\ \langle j_\mu(x)\rangle &= \tfrac{1}{2}e\langle[\bar{\psi}(x), \gamma_\mu\psi(x)]\rangle_0 . \end{aligned} \tag{A1.1}$$

In order to evaluate Eq. (A1.1), we introduce the electron Green's function $G(x, x')$,

$$G(x, x') = i\,\langle T(\psi(x)\,\bar{\psi}(x'))\rangle_0 \tag{A1.2}$$

which satisfies the differential equation

$$\left[m - \gamma^\mu\left(i\frac{\partial}{\partial x^\mu} - eA_\mu(x)\right)\right] G(x, x') = \delta^4(x - x'). \tag{A1.3}$$

It is easy to see that $\langle j_\mu(x)\rangle$ is just the limit of the Green's function in which x symmetrically approaches x',

$$\begin{aligned} ie\,\mathrm{Tr}[\gamma_\mu G(x, x)] &\equiv \tfrac{1}{2}ie\,\mathrm{Tr}[\gamma_\mu G(x, x')|_{x_0=x_0'+\epsilon} + \gamma_\mu G(x, x')|_{x_0=x_0'-\epsilon}] \\ &= \tfrac{1}{2}ie\,\mathrm{Tr}[\gamma_\mu i\langle\psi(x)\;\bar{\psi}(x)\rangle_0 - \gamma_\mu i\langle\bar{\psi}(x)\;\psi(x)\rangle_0^{\text{Dirac index transpose}}] \\ &= \tfrac{1}{2}e\langle[\bar{\psi}(x), \gamma_\mu\psi(x)]\rangle_0 = \langle j_\mu(x)\rangle. \end{aligned} \tag{A1.4}$$

Thus, to calculate the induced current it suffices to calculate the Green's function $G(x, x')$.

As Schwinger shows, this calculation is facilitated if we introduce a condensed notation in which $G(x, x')$ is regarded as the $(x \mid \cdots \mid x')$ matrix element of an operator G,

$$\begin{aligned} G(x, x') &\equiv (x \mid G \mid x'), \\ \delta(x - x') &\equiv (x \mid x'). \end{aligned} \tag{A1.5}$$

Introducing the additional operator π_μ,

$$\pi_\mu \equiv i\frac{\partial}{\partial x^\mu} - eA_\mu(x), \tag{A1.6}$$

the differential equation (A1.3) for the Green's function can be rewritten as the algebraic operator equation

$$(m - \gamma\cdot\pi)\,G = 1. \tag{A1.7}$$

Inverting Eq. (A1.7) and substituting into Eq. (A1.4), we get

$$\begin{aligned}
&\langle j_\mu(x)\rangle \\
&= ie\,\mathrm{Tr}\left[\gamma_\mu\left(x\left|\frac{1}{m-\gamma\cdot\pi}\right|x\right)\right] \\
&= \frac{1}{2}ie\,\mathrm{Tr}\left[\gamma_\mu\left(x\left|(m+\gamma\cdot\pi)\frac{1}{m^2-(\gamma\cdot\pi)^2}+\frac{1}{m^2-(\gamma\cdot\pi)^2}(m+\gamma\cdot\pi)\right|x\right)\right] \\
&= \frac{1}{2}ie\,\mathrm{Tr}\left[\gamma_\mu\left(x\left|\gamma\cdot\pi\frac{1}{m^2-(\gamma\cdot\pi)^2}+\frac{1}{m^2-(\gamma\cdot\pi)^2}\gamma\cdot\pi\right|x\right)\right],
\end{aligned} \tag{A1.8}$$

where we have used the fact that the trace of an odd number of γ matrices is zero. The next step is to exponentiate the operator $[m^2 - (\gamma\cdot\pi)^2]^{-1}$ using the identity

$$\begin{gathered}
\frac{1}{m^2-(\gamma\cdot\pi)^2} = i\int_0^\infty ds\; e^{-i[m^2-(\gamma\cdot\pi)^2]s} = i\int_0^\infty ds\; e^{-im^2s}U(s), \\
U(s) = e^{i(\gamma\cdot\pi)^2 s},
\end{gathered} \tag{A1.9}$$

which on substitution into Eq. (A1.8) gives

$$\langle j_\mu(x)\rangle = -\tfrac{1}{2}e\int_0^\infty ds\; e^{-im^2s}\,\mathrm{Tr}[\gamma_\mu\gamma_\nu(x\mid\pi^\nu U(s)\mid x) + \gamma_\nu\gamma_\mu(x\mid U(s)\,\pi^\nu\mid x)]. \tag{A1.10}$$

Let us now exploit the fact that $U(s)$ is just the "proper-time evolution" operator $U(s) = \exp[-i\mathscr{H}s]$ for the quantum mechanics problem with Hamiltonian

$$\begin{gathered}
\mathscr{H} = -(\gamma\cdot\pi)^2 = -\pi^2 - \tfrac{1}{2}e\sigma\cdot F, \qquad \sigma\cdot F \equiv \sigma_{\mu\nu}F^{\mu\nu}, \\
\sigma_{\mu\nu} = \tfrac{1}{2}i[\gamma_\mu, \gamma_\nu], \qquad F_{\mu\nu}(x) = \partial_\nu A_\mu - \partial_\mu A_\nu,
\end{gathered} \tag{A1.11}$$

and with "proper time" s. Introducing the definitions

$$\begin{gathered}
\mid x(0)) \equiv \mid x),\ (x(s)\mid \equiv (x(0)\mid U(s), \\
\pi(0) \equiv \pi, \qquad x(0) \equiv x, \\
\pi(s) \equiv U^{-1}(s)\,\pi(0)\,U(s),\ x(s) \equiv U^{-1}(s)\,x(0)\,U(s),
\end{gathered} \tag{A1.12}$$

we can rewrite Eq. (A1.10) as

$$\langle j_\mu(x)\rangle = -\tfrac{1}{2}e\int_0^\infty ds\, e^{-im^2 s}\,\mathrm{Tr}[\gamma_\mu\gamma_\nu(x(s)\mid \pi^\nu(s)\mid x(0)) + \gamma_\nu\gamma_\mu(x(s)\mid \pi^\nu(0)\mid x(0))], \tag{A1.13}$$

or equivalently, as

$$\langle j_\mu(x)\rangle = -\tfrac{1}{2}e\int_0^\infty ds\, e^{-im^2 s}\,\mathrm{Tr}[(x(s)\mid \pi_\mu(s) + \pi_\mu(0)\mid x(0)) - i\sigma_\mu{}^\nu(x(s)\mid \pi_\nu(s) - \pi_\nu(0)\mid x(0))], \tag{A1.14}$$

which is our final expression for the induced current. Using the commutation relations

$$[x_\mu\,,\,\pi_\nu] = -ig_{\mu\nu}\,,\qquad [\pi_\mu\,,\,\pi_\nu] = ieF_{\mu\nu}(x), \tag{A1.15}$$

we deduce from Eq. (A1.12) the following equations of motion satisfied by $x_\mu(s)$ and $\pi_\mu(s)$,

$$\begin{aligned}\frac{dx_\mu(s)}{ds} &= U^{-1}(s)[i\mathscr{H},\,x_\mu]\,U(s) = 2\pi_\mu(s),\\ \frac{d\pi_\mu(s)}{ds} &= U^{-1}(s)[i\mathscr{H},\,\pi_\mu]\,U(s)\\ &= -e[\pi^\nu(s)\,F_{\mu\nu}(x(s)) + F_{\mu\nu}(x(s))\,\pi^\nu(s) + \tfrac{1}{2}\Sigma_{\alpha\beta}(s)\,F^{\alpha\beta}_{.\mu}(x(s))],\end{aligned} \tag{A1.16}$$

with

$$F^{\alpha\beta}_{.\mu}(x) \equiv \partial F^{\alpha\beta}(x)/\partial x^u,\qquad \Sigma_{\alpha\beta}(s) \equiv U^{-1}(s)\,\sigma_{\alpha\beta}U(s). \tag{A1.17}$$

So far we have made no assumptions about the nature of the external field $F_{\mu\nu}(x)$. Now, let us specialize to the case in which $F_{\mu\nu}$ is the superposition of a strong constant magnetic field $\bar{F}$ and a plane wave of amplitude a and field strength f,

$$F_{\mu\nu}(x) = \bar{F}_{\mu\nu} + f_{\mu\nu}(\xi),\ \xi = n\cdot x, \tag{A1.18}$$

with $n = (1, \hat{n})$ a null vector defining the direction of propagation of the wave. Taking our constant magnetic field to point along the 3 axis, we have

$$\bar{F}_{21} = -\bar{F}_{12} = \bar{B};\ \text{all other components} = 0. \tag{A1.19}$$

To calculate the refractive indices, we will take

$$a_\mu(\xi) = \epsilon_\mu e^{i\omega\xi},$$
$$f_{\mu\nu}(\xi) = i\omega(n_\nu\epsilon_\mu - n_\mu\epsilon_\nu)\, e^{i\omega\xi}, \tag{A1.20}$$

and compute the term in $\langle j_\mu(x)\rangle$ linear in ϵ, while to calculate the photon splitting matrix element, we will take

$$a_\mu(\xi) = \epsilon_{1\mu} e^{i\omega_1\xi} + \epsilon_{2\mu} e^{i\omega_2\xi},$$
$$f_{\mu\nu}(\xi) = i\omega_1(n_\nu\epsilon_{1\mu} - n_\mu\epsilon_{1\nu})\, e^{i\omega_1\xi} + i\omega_2(n_\nu\epsilon_{2\mu} - n_\mu\epsilon_{2\nu})\, e^{i\omega_2\xi}, \tag{A1.21}$$

and compute the term in $\langle j_\mu(x)\rangle$ bilinear in ϵ_1 and ϵ_2.

The procedure now is as follows. First, we substitute Eqs. (A1.18) and (A1.19) into the equations of motion, Eq. (A1.16). We then perform two integrations which give us expressions for $\pi_\mu(s) \pm \pi_\mu(0)$ as linear combinations of a quantity Φ_μ, which is entirely of first order in the plane wave field $f_{\mu\nu}$, and of $x_\mu(s) - x_\mu(0)$:

$$\Phi_\mu(s) = -e[\pi^\nu(s) f_{\mu\nu}(\xi(s)) + f_{\mu\nu}(\xi(s))\, \pi^\nu(s) + \tfrac{1}{2}\Sigma_{\alpha\beta}(s) f^{\alpha\beta}_{,\mu}(\xi(s))],$$

$$\pi_\mu(s) - \pi_\mu(0) = C_\mu^{(-)\nu}[x_\nu(s) - x_\nu(0)] + \int_0^s dt\, \Phi_\mu(t),$$

$$\pi_\mu(s) + \pi_\mu(0) = C_\mu^{(+)\nu}[x_\nu(s) - x_\nu(0)] + \int_0^s dt\, T(s,t)_\mu{}^\nu\, \Phi_\nu(t),$$

$$C^{(\mp)} = c\text{-no. matrices}, \tag{A1.22}$$

$$T(s,t)_\mu{}^\nu = \begin{bmatrix} v & 0 & 0 & \\ 0 & C(s,t) & -S(s,t) & 0 \\ 0 & S(s,t) & C(s,t) & 0 \\ 0 & 0 & 0 & v \end{bmatrix} = \begin{bmatrix} T_0{}^0 & T_0{}^1 & \cdots & T_0{}^8 \\ T_1{}^0 & & & \vdots \\ \vdots & & & \vdots \\ T_8{}^0 & \cdots & & T_8{}^8 \end{bmatrix},$$

$$v = 2t/s - 1,$$

$$C(s,t) = \frac{\sin(e\bar{B}sv)}{\sin(e\bar{B}s)}, \qquad S(s,t) = \frac{\cos(e\bar{B}sv) - \cos(e\bar{B}s)}{\sin(e\bar{B}s)}.$$

Substituting Eq. (A1.22) into Eq. (A1.14), we find that the terms proportional to $x_\nu(s) - x_\nu(0)$ vanish, since $(x(s)|\, x_\nu(s) - x_\nu(0)|\, x(0)) = x_\nu - x_\nu = 0$, giving

$$\langle j_\mu(x)\rangle = -\tfrac{1}{2}e \int_0^\infty ds\, e^{-im^2 s} \int_0^s dt\, \mathrm{Tr}\{[T(s,t)_\mu{}^\nu - i\sigma_\mu{}^\nu](x(s)\mid \Phi_\nu(t)|\, x(0))\}. \tag{A1.23}$$

Our next step is to systematically develop Eq. (A1.23) in a power series in the plane wave amplitude. This is done by going over to an interaction picture in

which the zeroth order approximation describes the constant magnetic field alone, with no plane waves present. Thus, we write

$$
\begin{gathered}
U^{(0)}(s) = e^{-i\mathscr{H}^{(0)}s}, \\
\mathscr{H}^{(0)} = -\pi^{(0)2} - \tfrac{1}{2}e\sigma\cdot\bar{F}, \\
[x_\mu^{(0)}, \pi_\nu^{(0)}] = -ig_{\mu\nu}, \\
\pi^{(0)}(0) \equiv \pi^{(0)}, \qquad x^{(0)}(0) \equiv x^{(0)}, \\
\pi^{(0)}(s) \equiv U^{(0)-1}(s)\,\pi^{(0)}(0)\,U^{(0)}(s), \qquad x^{(0)}(s) \equiv U^{(0)-1}(s)\,x^{(0)}(0)\,U^{(0)}(s).
\end{gathered}
\tag{A1.24}
$$

As Schwinger has shown, the proper time evolution problem defined by Eq. (A1.24) can be simply and exactly integrated, giving

$$
\begin{aligned}
\pi_\mu^{(0)}(s) &= R(s)_\mu{}^\nu\,\pi_\nu^{(0)}(0), \\
x_\mu^{(0)}(s) &= x_\mu^{(0)}(0) + I(s)_\mu{}^\nu\,\pi_\nu^{(0)}(0), \\
R(s)_\mu{}^\nu &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(2e\bar{B}s) & -\sin(2e\bar{B}s) & 0 \\ 0 & \sin(2e\bar{B}s) & \cos(2e\bar{B}s) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\
I(s)_\mu{}^\nu &= \begin{bmatrix} 2s & 0 & 0 & 0 \\ 0 & \dfrac{\sin(2e\bar{B}s)}{e\bar{B}} & \dfrac{\cos(2e\bar{B}s)-1}{e\bar{B}} & 0 \\ 0 & \dfrac{1-\cos(2e\bar{B}s)}{e\bar{B}} & \dfrac{\sin(2e\bar{B}s)}{e\bar{B}} & 0 \\ 0 & 0 & 0 & 2s \end{bmatrix}, \\
(x^{(0)}(s)\mid x^{(0)}(0)) &= -i(4\pi)^{-2}\,\frac{es\bar{B}}{\sin(es\bar{B})}\,s^{-2}e^{\frac{1}{2}ies\sigma\cdot\bar{F}}.
\end{aligned}
\tag{A1.25}
$$

We now develop the problem in the presence of the plane wave field in a perturbation expansion around the zeroth order solution of Eq. (A1.25). At zero proper time, the exact and zeroth order coordinate are the same, while the exact and zeroth order canonical momenta differ just by the plane wave amplitude,

$$
x_\mu(0) = x_\mu^{(0)}(0), \qquad \pi_\mu(0) = \pi_\mu^{(0)}(0) - ea_\mu(\xi(0)) = \pi_\mu^{(0)}(0) - ea_\mu(\xi^{(0)}(0)). \tag{A1.26}
$$

To find the relation between the exact and zeroth order time evolution operators, we use Eq. (A1.26) to write

$$
\begin{aligned}
\mathscr{H} = \mathscr{H}^{(0)} + e[&a_\mu(\xi^{(0)})\,\pi^{(0)\mu} + \pi^{(0)\mu}a_\mu(\xi^{(0)}) \\
&- ea_\mu(\xi^{(0)})\,a^\mu(\xi^{(0)}) - \tfrac{1}{2}\sigma_{\alpha\beta}f^{\alpha\beta}(\xi^{(0)})].
\end{aligned}
\tag{A1.27}
$$

Defining $A = -i\mathscr{H}^{(0)}s$, $B = -i(\mathscr{H} - \mathscr{H}^{(0)})\, s$, we apply the identity

$$e^{A+B} = e^{A} T \exp\left[\int_0^1 dt\, e^{-At} B e^{At}\right], \tag{A1.28}$$

where T is the time ordering operation. Making a change of variable $u = st$, this gives

$$U(s) = U^{(0)}(s)\, U_I(s),$$

$$U_I(s) = T \exp\left\{-ie \int_0^s du\, [a_\mu(\xi^{(0)}(u))\, \pi^{(0)\mu}(u) + \pi^{(0)\mu}(u)\, a_\mu(\xi^{(0)}(u)) - e a_\mu(\xi^{(0)}(u))\, a^\mu(\xi^{(0)}(u)) - \tfrac{1}{2}\Sigma^{(0)}_{\alpha\beta}(u)\, f^{\alpha\beta}(\xi^{(0)}(u))]\right\}, \tag{A1.29}$$

where $\Sigma^{(0)}_{\alpha\beta}(u)$ is defined by

$$\Sigma^{(0)}_{\alpha\beta}(u) \equiv U^{(0)-1}(u)\, \sigma_{\alpha\beta} U^{(0)}(u) = e^{-\frac{1}{2}ieu\sigma\cdot F} \sigma_{\alpha\beta} e^{\frac{1}{2}ieu\sigma\cdot F}. \tag{A1.30}$$

As expected, the time evolution operator in the interaction picture is constructed from dynamical variables $\xi^{(0)}(u)$, $\pi^{(0)}(u)$ and $\Sigma^{(0)}_{\alpha\beta}(u)$ which have the proper-time dependence of the unperturbed problem.

We now use Eq. (A1.29) to rewrite our expression for $\langle j_\mu(x)\rangle$ in final form. Referring back to Eq. (A1.23), we write

$$\begin{aligned}(x(s) \mid \Phi_\nu(t) \mid x(0)) &= (x(0) \mid U(s)\, U^{-1}(t)\, \Phi_\nu(0)\, U(t) \mid x(0)) \\ &= (x(0) \mid U^{(0)}(s)\, U_I(s)\, U_I^{-1}(t)\, U^{(0)-1}(t)\, \Phi_\nu(0)\, U^{(0)}(t)\, U_I(t) \mid x(0)).\end{aligned} \tag{A1.31}$$

Making use of the explicit expression for Φ_ν in Eq. (A1.22), recalling Eq. (A1.26) and substituting Eq. (A1.31) back into Eq. (A1.23), we get

$$\begin{aligned}\langle j_\mu(x)\rangle = \tfrac{1}{2}e^2 \int_0^\infty ds\, e^{-im^2 s} \int_0^s dt\, \mathrm{Tr}\{[T(s,t)_\mu{}^\nu - i\sigma_\mu{}^\nu] \\ \times (x^{(0)}(s) \mid U_I(s)\, U_I^{-1}(t)[\pi^{(0)\nu}(t)\, f_{\mu\nu}(\xi^{(0)}(t)) + f_{\mu\nu}(\xi^{(0)}(t))\, \pi^{(0)\nu}(t) \\ - 2ea^\nu(\xi^{(0)}(t))\, f_{\mu\nu}(\xi^{(0)}(t)) + \tfrac{1}{2}\Sigma^{(0)}_{\alpha\beta}(t)\, f^{\alpha\beta}_{\cdot\mu}(\xi^{(0)}(t))]\, U_I(t) \mid x^{(0)}(0))\}.\end{aligned} \tag{A1.32}$$

By expanding the operators U_I in this equation to the requisite order in the plane

wave amplitude, we can obtain expressions for both the refractive indices and the photon splitting matrix element, as follows:

(i) *Refractive indices.* We take the plane wave amplitude as in Eq. (A1.20) and keep only first order terms in Eq. (A1.32), giving

$$\langle j_\mu(x)\rangle = \tfrac{1}{2}e^2 \int_0^\infty ds\, e^{-im^2 s} \int_0^s dt\, \mathrm{Tr}\{[T(s,t)_\mu{}^\nu - i\sigma_\mu{}^\nu](x^{(0)}(s) \mid \pi^{(0)\nu}(t) f_{\mu\nu}(\xi^{(0)}(t))$$

$$+ f_{\mu\nu}(\xi^{(0)}(t))\, \pi^{(0)\nu}(t) + \tfrac{1}{2}\Sigma^{(0)}_{\alpha\beta}(t) f^{\alpha\beta}_{,\mu}(\xi^{(0)}(t)) \mid x^{(0)}(0))\}. \qquad \text{(A1.33)}$$

To evaluate the matrix element in Eq. (A1.33), we use Eqs. (A1.25) to express $\pi^{(0)}(t)$ and $\exp[i\omega\xi^{(0)}(t)]$ in terms of $x^{(0)}(s)$ and $x^{(0)}(0)$,

$$\begin{aligned} \pi^{(0)}(t) &= R(t)\cdot I^{-1}(s)\cdot [x^{(0)}(s) - x^{(0)}(0)], \\ \exp[i\omega\xi^{(0)}(t)] &= \exp[i\omega n\cdot x^{(0)}(t)] \\ &= \exp\{i\omega[n\cdot(1 - I(t)\cdot I^{-1}(s))\cdot x^{(0)}(0) \\ &\quad + n\cdot I(t)\cdot I(s)^{-1}\cdot x^{(0)}(s)]\}. \end{aligned} \qquad \text{(A1.34)}$$

Since the commutator

$$[x^{(0)}_\mu(s), x^{(0)}_\nu(0)] = iI(s)_{\mu\nu} \qquad \text{(A1.35)}$$

is a c-number, we can then use the identities

$$\begin{aligned} e^a e^b &= e^b e^a e^{[a,b]} \\ e^a b &= b e^a + [a,b]\, e^a \end{aligned} \qquad \text{(A1.36)}$$

(valid when $[a,[a,b]] = [b,[a,b]] = 0$) to bring all factors $x^{(0)}(s)$ to the left and all factors $x^{(0)}(0)$ to the right in Eq. (A1.33), where they act on the left- and right-hand states to give c-numbers,

$$\begin{aligned} (x^{(0)}(s) \mid x^{(0)}(s) &= (x^{(0)}(s) \mid x, \\ x^{(0)}(0) \mid x^{(0)}(0)) &= x \mid x^{(0)}(0)). \end{aligned} \qquad \text{(A1.37)}$$

This leaves a completely c-number expression multiplied by the transformation function $(x^{(0)}(s) \mid x^{(0)}(0))$, which is given by Eq. (A1.25). Note that the matrix

element $(x^{(0)}(s)|\, \Sigma_{\alpha\beta}^{(0)}(t)|\, x^{(0)}(0))$ is equal to $(x^{(0)}(s)|\, x^{(0)}(0))\, \Sigma_{\alpha\beta}^{(0)}(t)$ but *not* to $\Sigma_{\alpha\beta}^{(0)}(t)(x^{(0)}(s)|\, x^{(0)}(0))$, since the state $|\, x^{(0)}(0))$ has no γ-matrix dependence, while the state $(x^{(0)}(s)|$ has the γ-matrix dependence of $U^{(0)}(s)$. Evaluating the γ-matrix traces, contracting tensor indices and making the change of integration variable (contour rotation) $s \to -is, t \to -it$ gives the final result. A particularly simple answer is obtained for the cases in which the plane wave is linearly polarized in the ($\|$) or ($\perp$) senses defined in the text. Taking, for simplicity, $\sin\theta = 1$, we get

$$\begin{aligned} &\| \text{ case:} && \langle j_\mu^{\|}\rangle = -\omega^2 A^{\|}[\omega, \bar{B}]\, a_\mu\,, \\ &\perp \text{ case:} && \langle j_\mu^{\perp}\rangle = -\omega^2 A^{\perp}[\omega, \bar{B}]\, a_\mu\,, \end{aligned} \tag{A1.38}$$

with $A^{\|,\perp}$ expressions given in Eq. (51) of the text.

To relate the refractive indices to $A^{\|,\perp}$, we note that in the self-consistent field approximation, the propagation eigenmodes satisfy the equation

$$\begin{aligned} \Box^2 a_\mu^{\|,\perp} &= \Box^2 \epsilon_\mu^{\|,\perp} e^{i\omega(x_0 - n_{\|,\perp}\hat{n}\cdot\mathbf{x})} = -\omega^2(1 - n_{\|,\perp}^2)\, a_\mu^{\|,\perp} \\ &= \langle j_\mu^{\|,\perp}\rangle \approx -\omega^2 A^{\|,\perp}[\omega, \bar{B}]\, a_\mu^{\|,\perp}. \end{aligned} \tag{A1.39}$$

This gives $n_{\|,\perp}^2 \approx 1 - A^{\|,\perp}[\omega, \bar{B}]$, or taking the square root,

$$n_{\|,\perp} \approx 1 - \tfrac{1}{2} A^{\|,\perp}[\omega, \bar{B}], \tag{A1.40}$$

as stated in the text. Note that in deriving Eq. (A1.40), we have in two places assumed that $n_{\|,\perp}$ are not much different from unity. The first place is in Eq. (A1.39), where we have used the coefficients $A^{\|,\perp}[\omega, \bar{B}]$ computed for plane waves satisfying the usual vacuum dispersion relation, rather than satisfying $k/\omega = n_{\|,\perp}$. (Recall that in Eq. (A1.18) we took $\xi = n \cdot x$, with n a null-vector.) The second place is, of course, in taking the square root to get Eq. (A1.40). Referring back to Eq. (55) of the text, we see that $n_{\|,\perp}$ will be close to unity provided that

$$\frac{\alpha}{\pi}\left(\frac{\bar{B}}{B_{CR}}\right)^2 \ll 1, \tag{A1.41}$$

a condition which is still well satisfied even when $\bar{B}/B_{CR}$ is of order unity. Our final formulas for the refractive indices are nearly identical with those obtained previously by Minguzzi [2], whose procedure we have followed rather closely. The only difference is that for the first term in $J^{\|,\perp}(s, v)$(see Eq. (51)) Minguzzi has

∓ 1 instead of $\mp e\bar{B}s \cosh(e\bar{B}sv)/\sinh(e\bar{B}s)$, an error which results[5] from his incorrectly replacing $\sum_{\alpha\beta}^{(0)}(t)$ in Eq. (A1.33) by its average $s^{-1}\int_0^s dt\, \sum_{\alpha\beta}^{(0)}(t)$.[6]

(ii) *Photon splitting matrix element.* To calculate the photon splitting matrix element, we take the plane wave amplitude as in Eq. (A1.21) and compute the second order terms in Eq. (A1.32), using the expression in Eq. (A1.29) for U_I. This gives

$$
\begin{aligned}
\langle j_\mu(x)\rangle = {}& \tfrac{1}{2}e^3 \int_0^\infty ds\, e^{-im^2 s} \int_0^s dt\, \mathrm{Tr}\{[T(s,t)_\mu{}^\nu - i\sigma_\mu{}^\nu] \\
& \times (x^{(0)}(s) \mid -2a^\nu(\xi^{(0)}(t)) f_{\mu\nu}(\xi^{(0)}(t)) \\
& - i\int_t^s du\, [a_\sigma(\xi^{(0)}(u))\, \pi^{(0)\sigma}(u) + \pi^{(0)\sigma}(u)\, a_\sigma(\xi^{(0)}(u)) - \tfrac{1}{2}\Sigma_{\delta\epsilon}^{(0)}(u) f^{\delta\epsilon}(\xi^{(0)}(u))] \\
& \times [\pi^{(0)\nu}(t) f_{\mu\nu}(\xi^{(0)}(t)) + f_{\mu\nu}(\xi^{(0)}(t))\, \pi^{(0)\nu}(t) + \tfrac{1}{2}\Sigma_{\alpha\beta}^{(0)}(t) f^{\alpha\beta}_{,\mu}(\xi^{(0)}(t))] \\
& - i[\pi^{(0)\nu}(t) f_{\mu\nu}(\xi^{(0)}(t)) + f_{\mu\nu}(\xi^{(0)}(t))\, \pi^{(0)\nu}(t) + \tfrac{1}{2}\Sigma_{\alpha\beta}^{(0)}(t) f^{\alpha\beta}_{,\mu}(\xi^{(0)}(t))] \\
& \times \int_0^t du\, [a_\sigma(\xi^{(0)}(u))\, \pi^{(0)\sigma}(u) + \pi^{(0)\sigma}(u)\, a_\sigma(\xi^{(0)}(u)) \\
& - \tfrac{1}{2}\Sigma_{\delta\epsilon}^{(0)}(u) f^{\delta\epsilon}(\xi^{(0)}(u))] \mid x^{(0)}(0)).
\end{aligned}
\tag{A1.42}
$$

[5] The error first appears in Minguzzi's analysis when, in his version of Eq. (AI. 16), he writes $\frac{1}{2}\sigma_{\alpha\beta} F^{\alpha\beta}_{,u}(s)$ instead of $\frac{1}{2}\Sigma_{\alpha\beta}(s) F^{\alpha\beta}_{,\mu}(s)$. This makes the final term of the matrix element in Eq. (AI. 33) read $(x^{(0)}(s) \mid \frac{1}{2}\sigma_{\alpha\beta} f^{\alpha\beta}_{,\mu}(\xi^{(0)}(t)) \mid x^{(0)}(0))$. Then, when evaluating $(x^{(0)}(s) \mid \sigma_{\alpha\beta} \mid x^{(0)}(0))$, instead of simply equating this to $(x^{(0)}(s) \mid x^{(0)}(0))\sigma_{\alpha\beta}$, Munguzzi notes that $(x^{(0)}(s) \mid x^{(0)}(0)) \propto \exp(\frac{1}{2}ies\sigma\cdot\bar{F})$ $\times$ (γ-matrix independent factors) and then regards $(x^{(0)}(s) \mid \sigma_{\alpha\beta} \mid x^{(0)}(0))$ as the variational derivative of $(x^{(0)}(s) \mid x^{(0)}(0))$ with respect to a small change in the constant field $\bar{F}$. Thus, he writes,

$$
\begin{aligned}
(x^{(0)}(s) \mid \sigma_{\alpha\beta} \mid x^{(0)}(0)) \text{ ``$=$'' } & \frac{\delta}{\delta\epsilon^{\alpha\beta}} (x^{(0)}(s) \mid x^{(0)}(0)) \Big|_{\frac{1}{2}iesF \to \frac{1}{2}iesF + \epsilon} \\
= {}& \frac{\delta}{\delta\epsilon^{\alpha\beta}} \exp(\tfrac{1}{2}ies\sigma\cdot\bar{F} + \epsilon\cdot\sigma) \times (\gamma\text{-matrix independent factors}) \\
= {}& (x^{(0)}(s) \mid x^{(0)}(0)) \frac{\delta}{\delta\epsilon^{\alpha\beta}}\, T\exp\left[\int_0^1 dt\, e^{-\frac{1}{2}iest\sigma\cdot F_\epsilon}\cdot\sigma\, e^{\frac{1}{2}iest\sigma\cdot F}\right] \\
= {}& (x^{(0)}(s) \mid x^{(0)}(0))\, s^{-1}\int_0^s dt\, \Sigma_{\alpha\beta}^{(0)}(t),
\end{aligned}
$$

where use has been made of Eqs. (AI. 28) and (AI. 30) and where, in the final line, the change of variable $st \to t$ has been made. So we see that Munguzzi's two errors result in his replacing $(x^{(0)}(s) \mid x^{(0)}(0))\, \Sigma_{\alpha\beta}^{(0)}(t)$ by the t-average of this quantity. As a result of this error, in Munguzzi's version of Eq. (51), the function $W^{\|}$ governing the convergence of the representation is $W^{\|} = \omega^2 t(1 - t/s)\text{-}m^2 s$, rather than the expression given in Eq. (53). This leads Munguzzi to the incorrect conclusion that absorptive contributions to $n_{\|}$ begin at $\omega = 2m$, rather than at the larger value $\omega = m[1 + (1 + 2\bar{B}/B_{CR})^{1/2}]$ obtained from Eq. (53). As we have noted, the larger value is the one which agrees with the parallel-photon photopair production threshold found by Toll.

[6] As we have implied in the text, the refractive index calculation is simplest when $\sin\theta = 1$. To obtain the answer for general $\sin\theta$, we note that the only Lorentz scalars on which the refractive

Letting $\langle j_\mu^{\text{bilinear}} \rangle$ denote the part of Eq. (A1.42) which is bilinear in ϵ_1 and ϵ_2, the matrix element $\mathscr{M}$ for photon splitting is

$$\mathscr{M} = -i(4\pi)^{3/2}\, \epsilon^\mu \langle j_\mu^{\text{bilinear}} \rangle, \tag{A1.43}$$

with ϵ the initial photon polarization. The evaluation of Eq. (A1.42) can be carried out by the same methods used to obtain the refractive indices, and leads to the result quoted in Eq. (25) of the text for the physically interesting case $(\|) \to (\perp)_1 + (\perp)_2$.[6]

APPENDIX II: Small Opening Angle Corrections to the Box Diagram

We estimate here the nonvanishing box diagram contribution to photon splitting which arises when spatial variation of the external magnetic field causes the three photon momenta to be nonparallel. With trivial modifications, as explained below, the calculation also applies to the case in which the external magnetic field is strictly constant and the nonparallelism results from vacuum dispersion effects. Our aim is to show that, in both of these cases, all terms in the box diagram matrix element are at least quadratic in the small angles ϕ_1, ϕ_2, ϕ_{12} between the photon wave vectors.

We proceed by considering the most general momentum-dependent term appearing in the part of the photon splitting matrix element which has one external field factor $\bar{F}$, when $\bar{F}$ carries nonvanishing four-momentum p. This is

$$FF^1F^2\bar{F}\ \underbrace{k \cdots k}_{\ell \text{ factors}}\ \underbrace{k_1 \cdots k_1}_{m \text{ factors}}\ \underbrace{k_2 \cdots k_2}_{n \text{ factors}}\ \underbrace{p \cdots p}_{r \text{ factors}}, \tag{A2.1}$$

with $\ell + m + n + r$ even and with all Lorentz indices contracted to form a Lorentz scalar. We consider various cases in turn:

(i) $r > 0$. Since $p = (0, \mathbf{p})$ and, according to Eq. (36) in the text, $|\mathbf{p}| \sim \phi^2$, Eq. (A2.1) is of order ϕ^2 at least.

(ii) $r = 0$. We distinguish three principal subcases.

(a) *Two photon four-momenta are contracted with $\bar{F}$.* Since the photon four-momenta are proportional, apart from terms of order ϕ, and since $\bar{F}$ is an

indices can depend are $\bar{F}_\mu{}^\nu \bar{F}_\nu{}^\mu = 2\bar{B}^2$ and $k^\mu \bar{F}_\mu{}^\nu \bar{F}_\nu{}^\eta k_\eta = \omega^2 \bar{B}^2 \sin^2\theta$, indicating that the recipe is simply to replace ω by $\omega \sin\theta$. A similar argument in the photon splitting case indicates that the matrix element for general θ is obtained from that for $\theta = \pi/2$ by making the replacements ω, ω_1, $\omega_2 \to \omega \sin\theta$, $\omega_1 \sin\theta$, $\omega_2 \sin\theta$.

VOLUME 77, NUMBER 9 PHYSICAL REVIEW LETTERS 26 AUGUST 1996

Photon Splitting in a Strong Magnetic Field: Recalculation and Comparison with Previous Calculations

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540

Christian Schubert
Humboldt Universitität zu Berlin, Invalidenstrasse 110, D-10115 Berlin, Germany
(Received 23 April 1996)

We recalculate the amplitude for photon splitting in a strong magnetic field below the pair production threshold, using the world line path integral variant of the Bern-Kosower formalism. Numerical comparison (using programs that we have made available for public access on the Internet) shows that the results of the recalculation are identical to the earlier calculations of Adler and later of Stoneham, and to the recent recalculation by Baier, Milstein, and Shaisultanov. [S0031-9007(96)01004-6]

PACS numbers: 12.20.Ds, 95.30.Cq

Photon splitting in a strong magnetic field is an interesting process, both from a theoretical viewpoint because of the relatively sophisticated methods needed to do the calculation, and because of its potential astrophysical applications. The first calculation to exactly include the corrections arising from nonzero photon frequency ω was given by Adler [1], who obtained the amplitude as a triple integral that is strongly convergent below the pair production threshold at $\omega = 2m$, and who included a numerical evaluation for the special case $\omega = m$. Subsequently, the calculation was repeated by Stoneham [2] using a different method, leading to a different expression as a triple integral, that has never been compared to the formula of Ref. [1] either analytically or numerically. Recently, a new calculation has been published by Mentzel, Berg, and Wunner [3] in the form of a triple infinite sum, and numerical evaluation of their formula by Wunner, Sang, and Berg [4] claims photon splitting rates roughly 4 orders of magnitude larger than those found in Ref. [1]. Since this result, if correct, would have important astrophysical implications, a recalculation by an independent method seems in order. We report the results of such a recalculation here, together with a numerical comparison of the resulting amplitude with those of Adler and of Stoneham, as well as with a recent recalculation independently carried out by Baier, Milstein, and Shaisultanov [5]. The comparison shows that these four independent calculations give precisely the same amplitude, showing no evidence of the dramatic energy dependent effects claimed in Refs. [3] and [4].

Our recalculation of the photon splitting amplitude uses a variant of the world line path integral approach to the Bern-Kosower formalism [6–9]. As is well known, the one loop QED effective action induced for the photon field by a spinor loop can be represented by the following double path integral:

$$\Gamma[A] = -2\int_0^\infty \frac{ds}{s} e^{-m^2 s} \int \mathcal{D}x\,\mathcal{D}\psi \exp\left[-\int_0^s d\tau\left(\frac{1}{4}\dot{x}^2 + \frac{1}{2}\psi\dot{\psi} + ieA_\mu\dot{x}^\mu - ie\psi^\mu F_{\mu\nu}\psi^\nu\right)\right]. \tag{1}$$

Here s is the usual Schwinger proper-time parameter, the $x^\mu(\tau)$'s are the periodic functions from the circle with circumference s into spacetime, and the $\psi^\mu(\tau)$'s are antiperiodic and Grassmann valued.

Photon scattering amplitudes are obtained by specializing the background to a sum of plane waves with definite polarizations. Both path integrals are then evaluated by one-dimensional perturbation theory; i.e., one obtains an integral representation for the N-photon amplitude by Wick-contracting N "photon vertex operators"

$$V = \int_0^T d\tau[\dot{x}^\mu\varepsilon_\mu - 2i\psi^\mu\psi^\nu k_\mu\varepsilon_\nu]\exp[ikx(\tau)]. \tag{2}$$

The appropriate one-dimensional propagators are

$$\begin{aligned}\langle y^\mu(\tau_1)y^\nu(\tau_2)\rangle &= -g^{\mu\nu}G_B(\tau_1,\tau_2) = -g^{\mu\nu}\left[\,|\,\tau_1-\tau_2\,| - \frac{(\tau_1-\tau_2)^2}{s}\right],\\ \langle\psi^\mu(\tau_1)\psi^\nu(\tau_2)\rangle &= \tfrac{1}{2}g^{\mu\nu}G_F(\tau_1,\tau_2) = \tfrac{1}{2}g^{\mu\nu}\mathrm{sign}(\tau_1-\tau_2).\end{aligned} \tag{3}$$

The bosonic Wick contraction is actually carried out in the relative coordinate $y(\tau) = x(\tau) - x_0$ of the closed loop, while the (ordinary) integration over the average position $x_0 = \frac{1}{s}\int_0^s d\tau\, x(\tau)$ yields energy-momentum conservation.

To take the additional constant magnetic background field B into account, one chooses Fock-Schwinger gauge, where its contribution to the world line Lagrangian becomes

$$\Delta\mathcal{L} = \tfrac{1}{2} iey^{\mu}F_{\mu\nu}\dot{y}^{\nu} - ie\psi^{\mu}F_{\mu\nu}\psi^{\nu}. \tag{4}$$

Being bilinear, those terms can be simply absorbed into the kinetic part of the Lagrangian [9,10]. This leads to generalized world line propagators defined by

$$\frac{1}{2}\left(\frac{\partial^2}{\partial\tau^2} - 2ieF\frac{\partial}{\partial\tau}\right)\mathcal{G}_B(\tau_1,\tau_2) = \delta(\tau_1-\tau_2) - \frac{1}{s}, \tag{5}$$

$$\frac{1}{2}\left(\frac{\partial}{\partial\tau} - 2ieF\right)\mathcal{G}_F(\tau_1,\tau_2) = \delta(\tau_1-\tau_2). \tag{6}$$

The solutions to these equations can be written in the form [11]

$$\mathcal{G}_B(\tau_1,\tau_2) = \frac{1}{2(eF)^2}\left(\frac{eF}{\sin(esF)}e^{-iesF\dot{G}_{B12}} + ieF\dot{G}_{B12} - \frac{1}{s}\right), \tag{7}$$

$$\mathcal{G}_F(\tau_1,\tau_2) = G_{F12}\frac{e^{-iesF\dot{G}_{B12}}}{\cos(esF)} \tag{8}$$

(we have abbreviated $G_{Bij} \equiv G_B(\tau_i,\tau_j)$, and a dot always denotes a derivative with respect to the first variable). Those expressions should be understood as power series in the field strength matrix. To obtain the photon splitting amplitude, we will use them for the Wick contraction of three vertex operators V_0 and $V_{1,2}$, representing the incoming and the two outgoing photons.

The calculation is greatly simplified by the special kinematics of this process. Energy-momentum conservation, $k_0 + k_1 + k_2 = 0$, forces collinearity of all three four-momenta, so that, writing $-k_0 \equiv k \equiv \omega n$,

$$k_1 = \frac{\omega_1}{\omega}k, \qquad k_2 = \frac{\omega_2}{\omega}k; \qquad k^2 = k_1^2 = k_2^2 = k\cdot k_1 = k\cdot k_2 = k_1\cdot k_2 = 0. \tag{9}$$

Moreover, a simple CP invariance argument together with an analysis of dispersive effects [1] shows that there is only one allowed polarization case. This is the one where the incoming photon is polarized parallel to the plane containing the external field and the direction of propagation, and both outgoing ones are polarized perpendicular to this plane. This choice of polarizations leads to the further vanishing relations

$$\varepsilon_{1,2}\cdot\varepsilon_0 = \varepsilon_{1,2}\cdot k = \varepsilon_{1,2}\cdot F = 0. \tag{10}$$

In particular, we cannot Lorentz contract ε_1 with anything but ε_2. This leaves us with only a small number of nonvanishing Wick contractions,

$$\langle V_0V_1V_2\rangle = \prod_{i=0}^{2}\int_0^T d\tau_i\, i\exp\left[\frac{1}{2}\sum_{i,j=0}^{2}\bar{\omega}_i\bar{\omega}_j n\mathcal{G}_{Bij}n\right]\Bigg\{[\varepsilon_1\ddot{\mathcal{G}}_{B12}\varepsilon_2 + \varepsilon_1\mathcal{G}_{F12}\varepsilon_2\bar{\omega}_1\bar{\omega}_2 n\mathcal{G}_{F12}n] \times\left[-\sum_{i=0}^{2}\bar{\omega}_i\varepsilon_0\dot{\mathcal{G}}_{B0i}n + \bar{\omega}_0\varepsilon_0\mathcal{G}_{F00}n\right] - \bar{\omega}_0\bar{\omega}_1\bar{\omega}_2\varepsilon_1\mathcal{G}_{F12}\varepsilon_2[n\mathcal{G}_{F10}\varepsilon_0\, n\mathcal{G}_{F20}n - (1\leftrightarrow 2)]\Bigg\}. \tag{11}$$

For compact notation we have defined $\bar{\omega}_0 = \omega, \bar{\omega}_{1,2} = -\omega_{1,2}$. This result has still to be multiplied by an overall factor of $(esB)\cosh(esB)/(4\pi s)^2\sinh(esB)$, which by itself would just produce the Euler-Heisenberg Lagrangian, and here appears as the product of the two free Gaussian path integrals [8].

It is then a matter of simple algebra to obtain the following representation for the matrix element $C_2[\omega,\omega_1,\omega_2,B]$ appearing in Eq. (25) of [1]:

$$\begin{aligned} C_2[\omega,\omega_1,\omega_2,B] = {}& \frac{m^8}{4\omega\,\omega_1\omega_2}\int_0^\infty ds\, s\,\frac{e^{-m^2 s}}{(esB)^2\sinh(esB)}\int_0^s d\tau_1\int_0^s d\tau_2 \\ &\times\exp\left\{-\frac{1}{2}\sum_{i,j=0}^{2}\bar{\omega}_i\bar{\omega}_j\left[G_{Bij} + \frac{1}{2eB}\frac{\cosh(esB\dot{G}_{Bij})}{\sinh(esB)}\right]\right\} \\ &\times\Bigg\{[-\cosh(esB)\ddot{G}_{B12} + \omega_1\omega_2(\cosh(esB) - \cosh(esB\dot{G}_{B12}))] \\ &\times\left[\omega(\coth(esB) - \tanh(esB)) - \omega_1\frac{\cosh(esB\dot{G}_{B01})}{\sinh(esB)} - \omega_2\frac{\cosh(esB\dot{G}_{B02})}{\sinh(esB)}\right] \\ &+ \omega\,\omega_1\omega_2\frac{G_{F12}}{\cosh(esB)}[\sinh(esB\dot{G}_{B01})(\cosh(esB) - \cosh(esB\dot{G}_{B02})) - (1\leftrightarrow 2)]\Bigg\}. \end{aligned} \tag{12}$$

Here translation invariance in τ has been used to set the position τ_0 of the incoming photon equal to s. Coincidence limits have to be treated according to the rules $\dot{G}_B(\tau,\tau) = 0, \dot{G}_B^2(\tau,\tau) = 1$.

VOLUME 77, NUMBER 9 PHYSICAL REVIEW LETTERS 26 AUGUST 1996

Alternatively, one may remove $\ddot{G}_{B12}$ by partial integration on the circle. This leads to the equivalent formula

$$
\begin{aligned}
C_2[\omega,\omega_1,\omega_2,B] = {} & \frac{m^8}{4}\int_0^\infty ds\, s\, e^{-m^2 s}\frac{\cosh(esB)}{(esB)^2\sinh(esB)}\int_0^s d\tau_1\int_0^s d\tau_2 \\
& \times \exp\biggl\{-\frac{1}{2}\sum_{i,j=0}^{2}\bar{\omega}_i\bar{\omega}_j\biggl[G_{Bij}+\frac{1}{2eB}\frac{\cosh(esB\dot{G}_{Bij})}{\sinh(esB)}\biggr]\biggr\} \\
& \times\biggl\{\biggl[\dot{G}_{B12}\biggl(\dot{G}_{B12}-\frac{\sinh(esB\dot{G}_{B12})}{\sinh(esB)}\biggr)-\biggl(1-\frac{\cosh(esB\dot{G}_{B12})}{\cosh(esB)}\biggr)\biggr] \\
& \times\biggl[-\coth(esB)+\tanh(esB)+\frac{\omega_1}{\omega}\frac{\cosh(esB\dot{G}_{B01})}{\sinh(esB)}+\frac{\omega_2}{\omega}\frac{\cosh(esB\dot{G}_{B02})}{\sinh(esB)}\biggr] \\
& +\dot{G}_{B12}\biggl[\biggl(\frac{\cosh(esB\dot{G}_{B02})}{\sinh(esB)}-\frac{1}{esB}\biggr)\biggl(\dot{G}_{B01}-\frac{\sinh(esB\dot{G}_{B01})}{\sinh(esB)}\biggr)-(1\leftrightarrow 2)\biggr] \\
& +\frac{1}{2}\dot{G}_{B12}\biggl[\frac{\omega}{\omega_2}\biggl(\dot{G}_{B01}-\frac{\sinh(esB\dot{G}_{B01})}{\sinh(esB)}\biggr)-(1\leftrightarrow 2)\biggr]\biggl(-\coth(esB)+\frac{1}{esB}+\tanh(esB)\biggr) \\
& +G_{F12}\biggl[\frac{\sinh(esB\dot{G}_{B01})}{\cosh(esB)}\biggl(1-\frac{\cosh(esB\dot{G}_{B02})}{\cosh(esB)}\biggr)-(1\leftrightarrow 2)\biggr]\biggr\}. \qquad (13)
\end{aligned}
$$

This form of the amplitude is less compact, but the integrand (apart from the exponential) is homogeneous in the ω_i.

Finally, let us remark that the analogous expression for scalar QED would be obtained by deleting all terms in Eq. (11) containing a $\mathcal{G}_F$, as well as the $\cosh(esB)$ appearing in the overall factor and the global factor of -2 in Eq. (1).

In order to compare the amplitudes of Eqs. (12) and (13) to those of Refs. [1], [2], and [5], we observe that both Eqs. (12) and (13) can be written in the form

$$
C_2[\omega,\omega_1,\omega_2,B] = \frac{m^8}{4B^2\omega\omega_1\omega_2} \times\int_0^\infty \frac{ds}{s}e^{-m^2 s}J_2(s,\omega,\omega_1,\omega_2,B), \qquad (14)
$$

in which J_2 is independent of the electron mass m. Inspection shows that the amplitude expressions of Adler [1] and Baier, Milstein, and Shaisultanov [5] are already in the form of Eq. (14), while that of Stoneham [2] can be put in this form by doing an integration by parts in the proper time parameter s, using the identity

$$
m^2 e^{-m^2 s} = -\frac{d}{ds}e^{-m^2 s} \qquad (15)
$$

to eliminate a term proportional to m^2 in the amplitude. In rewriting Stoneham's formulas in this form, we note that his $M_1(B)$ is what we are calling $C_2[\omega,\omega_1,\omega_2,B]$, and that there is an error of an overall minus sign in either his Eq. (37) or the first line of his Eq. (40). Similarly, in rewriting the formulas of Baier, Milstein, and Shaisultanov in this form, we note that their amplitude T is related to C_2 by

$$
C_2[\omega,\omega_1,\omega_2,B] = \frac{\pi^{1/2}m^8}{4\alpha^3B^3\omega\omega_1\omega_2}T. \qquad (16)
$$

Once all amplitudes are put in the form of Eq. (14), we can compare them by comparing the proper time integrand $J_2(s,\omega,\omega_1,\omega_2,B)$, which in each case involves only a double integral over a bounded domain. The only remaining subtlety is that we must remember that J_2 vanishes as $\omega\omega_1\omega_2$ for small photon energy; this is manifest in Eq. (13) above, but in Eq. (12) and the corresponding equations obtained from Refs. [1], [2], and [5], there is an apparent linear term in the frequencies which vanishes when the double integral is done exactly. In order to get robust results for small photon frequency when the double integral is done numerically, this linear term must first be subtracted away, by replacing expressions of the form

$$
\int\int e^Q(L+C), \qquad (17a)
$$

with L, Q, and C, respectively, linear, quadratic, and cubic in the photon frequencies, by the subtracted expression

$$
\int\int[(e^Q-1)L+e^QC]. \qquad (17b)
$$

This subtraction is already present in the expression of Eq. (25) of Ref. [1], and is discussed in the form of Eqs. (17a) and (17b) in Ref. [5], and it also must be applied to Eqs. (37) and (39) of Ref. [2] after the integration by parts of Eq. (15) has been carried out. While in principle this subtraction should be applied to Eq. (12) above, it turns out not to be needed there, because the linear term in the frequencies involves only integrals of the general form

$$
\int_0^s d\tau_1 f(s,\tau_1)\int_0^s d\tau_2[\delta(\tau_1-\tau_2)-1/s], \qquad (18)
$$

which is exactly zero using a discrete center-of-bin integration method when the δ function is discretized as a Kronecker delta. Thus Eq. (12) is robust for small photon

frequencies as it stands, when used in conjunction with center-of-bin integration.

With these preliminaries out of the way, it is then completely straightforward to program the functions $J_2(s, \omega, \omega_1, \omega_2, B)$ for the five cases represented by the formulas of Adler [2], Stoneham [3], Eq. (12) of this paper, Eq. (13) of this paper, and Baier, Milstein, and Shaisultanov [5], with the result that they are all seen to be precisely the same; the residual errors approach zero quadratically as the integration mesh spacing approaches zero, as expected for trapezoidal integration. We have not carried out the s and ω_1 integrals needed to get the photon splitting absorption coefficient, since this was done in Ref. [1], with results confirmed by the more extensive numerical analysis given in Ref. [5]. However, anyone wishing to do this further computation can obtain our programs for calculating the proper time integrand J_2 by accessing S. L. A.'s home page at the Institute for Advanced Study [12].

C. S. would like to thank P. Haberl for help with numerical work and the DFG for financial support. S. L. A. wishes to acknowledge the hospitality of the Institute for Theoretical Physics in Santa Barbara, where parts of this work were done. He also wishes to thank J. N. Bahcall for suggesting that this work be undertaken and V. N. Baier for informing him of the results of Ref. [5]. This work was supported in part by the Department of Energy under Grant No. DE-FG02-90ER40542.

[1] S. L. Adler, Ann. Phys. (N.Y.) **67**, 599 (1971).
[2] R. J. Stoneham, J. Phys. A **12**, 2187 (1979).
[3] M. Mentzel, D. Berg, and G. Wunner, Phys. Rev. D **50**, 1125 (1994).
[4] G. Wunner, R. Sang, and D. Berg, Astrophys. J. **455**, L51 (1995).
[5] V. N. Baier, A. I. Milstein, and R. Zh. Shaisultanov, preceding Letter, Phys. Rev. Lett. **77**, 1691 (1996).
[6] Z. Bern and D. A. Kosower, Phys. Rev. Lett. **66**, 1669 (1991); Nucl. Phys. **B379**, 451 (1992).
[7] M. J. Strassler, Nucl. Phys. **B385**, 145 (1992).
[8] M. G. Schmidt and C. Schubert, Phys. Lett. B **318**, 438 (1993).
[9] D. Cangemi, E. D'Hoker, and G. Dunne, Phys. Rev. D **51**, 2513 (1995).
[10] D. C. G. McKeon and T. N. Sherry, Mod. Phys. Lett. A **9**, 2167 (1994).
[11] M. Reuter, M. G. Schmidt, and C. Schubert (to be published).
[12] Institute for Advanced Study web site: http://www.sns.ias.edu/~adler/Html/photonsplit.html

PHYSICAL REVIEW D VOLUME 4, NUMBER 10 15 NOVEMBER 1971

Quantum Electrodynamics Without Photon Self-Energy Parts: An Application of the Callan-Symanzik Scaling Equations*

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540

and

William A. Bardeen
Institute of Theoretical Physics, Department of Physics, Stanford University, Stanford, California 94305

(Received 16 August 1971)

In a series of recent papers, Johnson, Baker, and Willey study quantum electrodynamics with internal photon self-energy parts omitted. They find that in this model the asymptotic electron and photon propagators have remarkable, simple properties. In the present note we show that these properties can be derived in a very economical fashion by using the Callan-Symanzik scaling equations.

In a series of recent papers, Johnson, Baker, and Willey[1] (JBW) have examined the question of whether quantum electrodynamics can be a self-consistent, finite theory. They start from the assumption that the Gell-Mann–Low eigenvalue condition[2] has a finite root e_0, giving the renormalized photon propagator the asymptotic behavior[3]

$$e^2\tilde{D}_F'(q)_{\mu\nu} \underset{q^2 \gg m^2}{\sim} \frac{-g_{\mu\nu}e_0^2}{q^2} + \text{gauge terms}, \tag{1}$$

m = electron mass,

e_0 = finite bare charge.

A simple application of Weinberg's theorem[4] then shows that the asymptotic behavior of the renormalized electron propagator $\tilde{S}_F'(p)$ is correctly obtained[5] by replacing all internal photon propagators by their asymptotic form, Eq. (1). Thus, one is led to consider quantum electrodynamics without internal photon self-energy parts. In this model, Baker and Johnson[1] find, using renormalization-group methods, that the asymptotic electron propagator has the remarkably simple form

$$\tilde{S}_F'(p)^{-1} \underset{p^2 \gg m^2}{\sim} C[\gamma\cdot p + am(m^2/-p^2)^\epsilon]. \tag{2}$$

Here ϵ is a power series in $\alpha_0 = e_0^2/(4\pi)$,

$$\epsilon = \frac{3}{2}\left(\frac{\alpha_0}{2\pi}\right) + \frac{3}{8}\left(\frac{\alpha_0}{2\pi}\right)^2 + \cdots, \tag{3}$$

a is a constant, and (in the Landau gauge where the electron wave-function renormalization Z_2 is finite) C is another constant. According to Eq. (2), if $\epsilon > 0$ [as suggested by the leading terms in the expansion of Eq. (3)], the asymptotic electron propagator is identical to the propagator of a free, massless fermion. This means that the electron bare mass m_0 is zero in the limit of infinite cutoff, and not divergent, as would be indicated by expanding $(m^2/-p^2)^\epsilon$ in a perturbation expansion in α_0 and truncating at a finite order.

Johnson, Baker, and Willey[1] have also studied the photon propagator in the model with no internal photon self-energy insertions. Introducing a cutoff Λ^2 to define the unrenormalized photon propagator and photon proper self-energy $D_F'(q)_{\mu\nu}$ and $\pi(q^2)$,[5]

$$e_0^2 D_F'(q)_{\mu\nu} = \frac{-g_{\mu\nu}e_0^2}{q^2[1+e_0^2\pi(q^2)]} + \text{gauge terms}, \tag{4}$$

they find that, asymptotically,

$$\pi(q^2) \underset{q^2 \gg m^2\ \Lambda^2 \gg m^2}{\sim} h(\alpha_0) + f(\alpha_0)\ln(-q^2/\Lambda^2). \tag{5}$$

For Eq. (5) to be consistent with the ansatz of Eq. (1) the logarithmically divergent term in Eq. (5) must vanish. This gives the simplified eigenvalue condition

$$f(\alpha_0) = 0, \tag{6}$$

which involves only vacuum-polarization graphs without internal photon self-energy parts. Equation (6) has been shown[1] to be equivalent to the Gell-Mann–Low eigenvalue condition (which involves all vacuum-polarization graphs), so the discussion starting from Eq. (1) is self-consistent.

The purpose of the present paper is to consider quantum electrodynamics without internal photon self-energy parts from the viewpoint of the Callan-Symanzik[6] scaling equations. The basic idea which we exploit is that when photon self-energy parts are omitted, the troublesome coupling-constant-derivative terms, which would destroy scaling behavior, do not appear in the Callan-Symanzik scaling equations for quantum electrodynamics.[7] As a result, application of the scaling equations in asymptotic situations leads to simple scaling be-

havior with an "anomalous" dimension. But this is just the type of behavior which Baker and Johnson find for the mass term in Eq. (2), so it is not surprising that the Callan-Symanzik equations lead to an economical derivation of Eq. (2). The same methods, we find, lead to a simple derivation of Eq. (5) as well.

In deriving the Callan-Symanzik equations, we follow closely a method due to Coleman.[8] We first make the unrenormalized quantities m_0, Z_2, and $\pi(q^2)$ finite by introducing an ultraviolet cutoff Λ^2 and an infrared cutoff μ^2 in the following manner:

(i) We take the propagator for internal photons to be

$$D_F^0(q)_{\mu\nu} = \left(\frac{\xi q_\mu q_\nu}{q^2} - g_{\mu\nu}\right)\frac{1}{q^2-\mu^2+i\epsilon}\,\frac{-\Lambda^2}{q^2-\Lambda^2+i\epsilon}\,. \tag{7}$$

This means that we are working in massive electrodynamics with photon mass μ^2. Since internal photon self-energy parts are omitted in our model, there is no distinction between bare and physical photon mass.

(ii) We calculate the lowest-order vacuum-polarization contribution to $\pi(q^2)$ [see Fig. 1(a)] in the following manner. First we impose gauge invariance to remove the quadratic divergence, and then we regulate the fermion loop, with fermion regulator mass Λ, to remove the logarithmic divergence.

(iii) All vacuum-polarization loops with four or more vertices [see Fig. 1(b)] are calculated by imposing gauge invariance, which makes them finite. The requirement of gauge-invariant calculation of loops, together with the photon-propagator cutoff specified in (i), renders convergent the vacuum-polarization contributions to $\pi(q^2)$ of the type illustrated in Fig. 1(c). As a result of this cutoff scheme, the quantities m_0, Z_2, and $\pi(q^2)$ become Λ-dependent. On the other hand, because we omit internal photon self-energy parts, the photon coupling constant e_0 is a fixed number, independent of Λ and of the physical electron and photon masses m and μ.

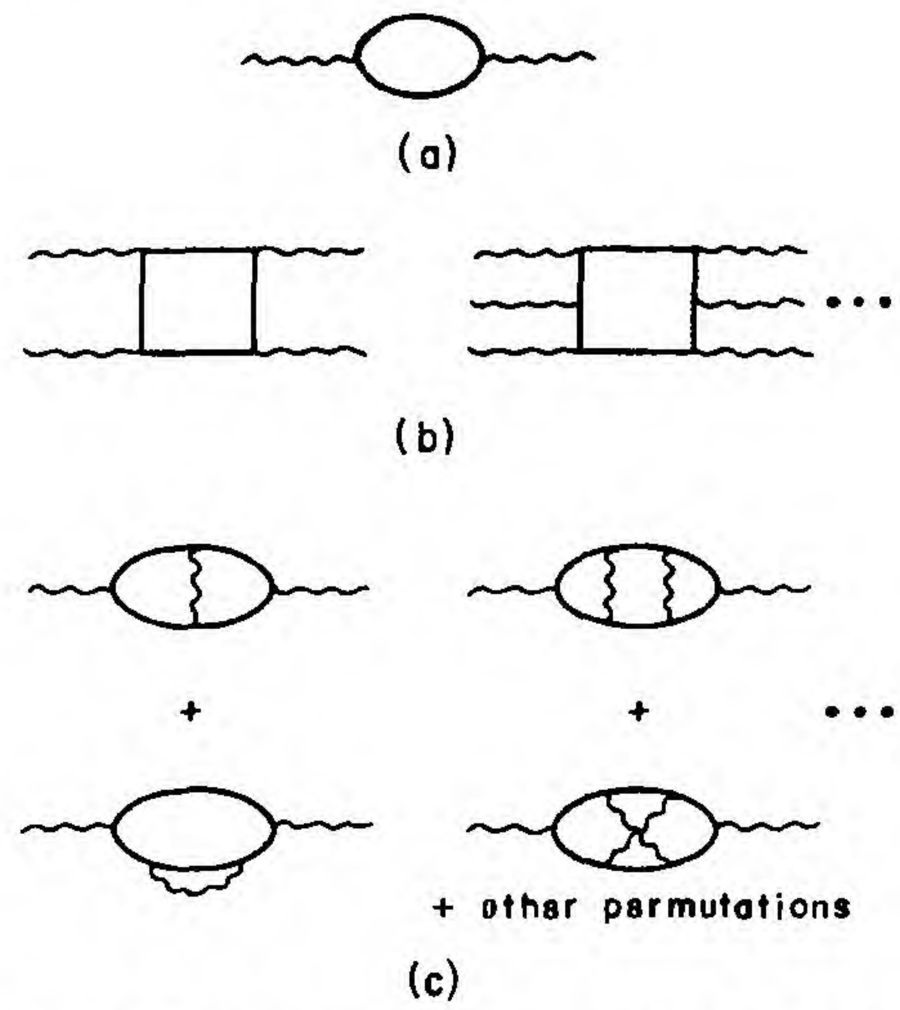


FIG. 1. (a) Lowest-order vacuum-polarization contribution to $\pi(q^2)$. (b) Vacuum-polarization loops with four or more vertices. (c) Vacuum-polarization contributions to $\pi(q^2)$ which involve the loops with four or more vertices illustrated in (b).

Having precisely specified our model, we are ready to discuss the scaling behavior of the electron propagator. The renormalized and unrenormalized electron propagators are related by the equation

$$\tilde{S}_F'(p)^{-1} = Z_2 S_F'(p)^{-1} = Z_2[\gamma\cdot p - m_0 - \Sigma(p)]\,, \tag{8}$$

with $\Sigma(p)$ the unrenormalized electron proper self-energy. Let us consider the change in Eq. (8) when the physical electron and photon masses m and μ are varied, with the ratio μ/m, with Λ, and with e_0 all held fixed. This is described by acting on Eq. (8) with the differential operator $m(\partial/\partial m) + \mu(\partial/\partial\mu)$, giving

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}\right)\tilde{S}_F'(p)^{-1} = \left[\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}\right)Z_2\right][\gamma\cdot p - m_0 - \Sigma(p)] - Z_2\left[\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}\right)m_0\right]\left[1+\frac{\partial\Sigma(p)}{\partial m_0}\right] - Z_2\mu\frac{\partial\Sigma(p)}{\partial\mu}\,. \tag{9}$$

On the right-hand side of Eq. (9) the operator $m(\partial/\partial m) + \mu(\partial/\partial\mu)$ is understood to act only on the quantity enclosed with it in square brackets; in deriving this equation, we have used the fact that $\Sigma(p)$ can depend on the physical mass m only through the bare mass m_0. The second and third terms on the right-hand side can be simply interpreted as follows: The quantity $1+\partial\Sigma(p)/\partial m_0$ appearing in the second term is just the zero-momentum-transfer vertex of the scalar electron current $j_S = \bar{\psi}\psi$,

$$1+\frac{\partial\Sigma(p)}{\partial m_0} = \Gamma_S(p,p)\,. \tag{10}$$

Typical diagrams contributing to $\Gamma_S(p,p)$ are illustrated in Fig. 2(a). Because internal photon self-

energy parts are omitted from $S_F'(p)^{-1}$, they are omitted from $\Gamma_S(p,p)$ as well; absent in addition are diagrams of the type shown in Fig. 2(b), which would arise from electron-mass differentiation of an internal photon self-energy part. The quantity $\mu\partial\Sigma(p)/\partial\mu$ appearing in the third term is a second type of scalar vertex at zero momentum transfer,

$$\mu\frac{\partial\Sigma(p)}{\partial\mu}\equiv\mu^2\Gamma_{S'}(p,p)\,. \tag{11}$$

Diagramatically, it is the sum of contributions obtained by replacing successively each internal photon propagator (of four-momentum, say, q) by

$$\text{photon propagator }(q)\times\frac{2\mu^2}{q^2-\mu^2+i\epsilon}\,, \tag{12}$$

as illustrated in Fig. 3.

The next step is to reexpress the right-hand side of Eq. (9) in terms of renormalized quantities. Since the skeleton graphs for $\Gamma_{S'}(p,p)$ are all convergent, this vertex is made finite by multiplication by Z_2,

$$Z_2\Gamma_{S'}(p,p)\equiv\tilde{\Gamma}_{S'}(p,p)$$
$$=\text{cutoff-independent as }\Lambda\to\infty\,. \tag{13}$$

By contrast, the vertex $\Gamma_S(p,p)$ has divergent skeleton graphs [see Fig. 2(a)], and so needs a vertex-renormalization factor in addition to the wave-function renormalization Z_2. In Appendix A we show that this factor is just the bare mass m_0,

$$m_0Z_2\Gamma_S(p,p)\equiv m\tilde{\Gamma}_S(p,p)$$
$$=\text{cutoff-independent as }\Lambda\to\infty\,. \tag{14}$$

Hence Eq. (9) takes the final form

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma\right)\tilde{S}_F'(p)^{-1}$$
$$=-m(1+\alpha)\tilde{\Gamma}_S(p,p)-\mu^2\tilde{\Gamma}_{S'}(p,p)\,, \tag{15a}$$

with

$$\gamma=-Z_2^{-1}\left[\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}\right)Z_2\right],$$
$$1+\alpha=m_0^{-1}\left[\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}\right)m_0\right]. \tag{15b}$$

Equation (15) is a typical Callan-Symanzik scaling equation, as simplified by the neglect of internal photon self-energy parts.[9]

Let us now consider the behavior of Eq. (15a) as p becomes infinite in a spacelike direction. We will keep all terms which are constant or which grow as powers of $\ln p^2$, but will drop terms which vanish as $(p^{-1},p^{-2},\ldots)\times$(powers of $\ln p^2$). By a simple application of Weinberg's theorem to the

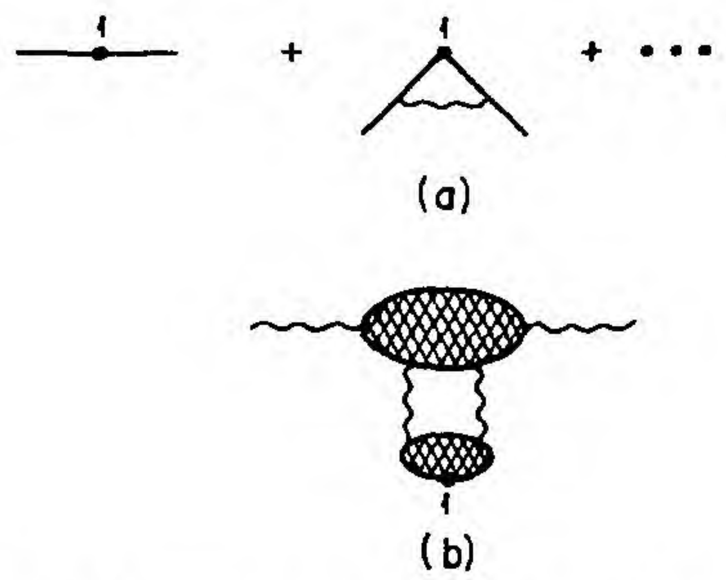


FIG. 2. (a) Typical diagrams contributing to $\Gamma_S(p,p)$. (b) Type of diagrams which are omitted from $\Gamma_S(p,p)$, because they could arise only from electron-mass differentiation of an internal photon self-energy part.

graphs which contribute to $\tilde{\Gamma}_S$ and $\tilde{\Gamma}_{S'}$ (see Figs. 2 and 3) we find that, to any finite order of perturbation theory,

$$\tilde{\Gamma}_S(p,p)\underset{p\to\infty}{\sim}(\text{powers of }\ln p^2)\,,$$
$$\tilde{\Gamma}_{S'}(p,p)\underset{p\to\infty}{\sim}p^{-1}\times(\text{powers of }\ln p^2)\,. \tag{16}$$

Thus, in the asymptotic limit, $\tilde{\Gamma}_S$ must be retained in Eq. (15) but $\tilde{\Gamma}_{S'}(p,p)$ may be dropped. Furthermore, introducing the general functional forms

$$\tilde{S}_F'(p)^{-1}=\gamma\cdot pF(p^2/m^2,\mu^2/m^2,e_0)$$
$$+mG(p^2/m^2,\mu^2/m^2,e_0)\,,$$
$$m\tilde{\Gamma}_S(p,p)=\gamma\cdot pH(p^2/m^2,\mu^2/m^2,e_0)$$
$$+mJ(p^2/m^2,\mu^2/m^2,e_0)\,, \tag{17}$$

and applying Weinberg's theorem again, we find that F, G, H, J have, to any finite order of perturbation theory, the asymptotic behavior

$$F,G,J\underset{p\to\infty}{\sim}(\text{powers of }\ln p^2)\,,$$
$$H\underset{p\to\infty}{\sim}p^{-1}\times(\text{powers of }\ln p^2)\,. \tag{18}$$

Substituting Eq. (17) into Eq. (15), equating separately the coefficients of $\gamma\cdot p$ and m, and dropping terms which vanish asymptotically, we get

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma\right)F(p^2/m^2,\mu^2/m^2,e_0)\underset{p\to\infty}{\sim}0\,, \tag{19a}$$

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma\right)mG(p^2/m^2,\mu^2/m^2,e_0)$$
$$\underset{p\to\infty}{\sim}-m(1+\alpha)J(p^2/m^2,\mu^2/m^2,e_0)\,. \tag{19b}$$

From Eq. (19a) we learn a number of things.

FIG. 3. A typical diagram contributing to $\Gamma_{S'}(p,p)$. The doubled photon propagator denotes $2ig_{\mu\nu}\Lambda^2\mu^2(q^2-\mu^2+i\epsilon)^{-2}(q^2-\Lambda^2+i\epsilon)^{-1}$.

First, since everything in this equation except γ is cutoff-independent, γ must be cutoff-independent,[10] i.e.,

$$\gamma=\gamma(\mu^2/m^2, e_0). \tag{20}$$

Comparing with Eq. (15a), we then learn that α is cutoff-independent also,[10] i.e.,

$$\alpha=\alpha(\mu^2/m^2, e_0). \tag{21}$$

[We will see shortly that there is actually no dependence on μ^2/m^2 in Eqs. (20) and (21).] Integrating Eq. (15b), we find that the Λ dependence of Z_2 and m_0 is given by

$$\begin{aligned} Z_2&=C_1(\mu^2/m^2, e_0)(\Lambda^2/m^2)^{\gamma/2}, \\ m_0&=C_2(\mu^2/m^2, e_0)m(\Lambda^2/m^2)^{-\alpha/2}, \end{aligned} \tag{22}$$

with C_1 and C_2 dimensionless functions. Finally, integrating Eq. (19a) we find that F has a power dependence on p^2 for large spacelike p,

$$F(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} f_1(\mu^2/m^2, e_0)(-p^2/m^2)^{\gamma/2}. \tag{23}$$

To obtain the equation satisfied by G which is analogous to Eq. (23), we must study the asymptotic behavior of the quantity J appearing on the right-hand side of Eq. (19b). We do this by calculating the Callan-Symanzik scaling equation satisfied by $\tilde{\Gamma}_S(p,p)$. Starting from Eq. (14), and proceeding in analogy with the calculation of Eqs. (9)-(15), we find

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma-\alpha\right)\tilde{\Gamma}_S(p,p) =m(1+\alpha)\tilde{T}_{SS}(p,p)+\mu^2\tilde{T}_{SS'}(p,p), \tag{24}$$

with the electron-scalar four-point functions $\tilde{T}_{SS}$ and $\tilde{T}_{SS'}$ defined by

$$\begin{aligned} m^2\tilde{T}_{SS}(p,p)&=m_0^2Z_2\frac{\partial}{\partial m_0}\Gamma_S(p,p), \\ m\mu^2\tilde{T}_{SS'}(p,p)&=m_0Z_2\mu\frac{\partial}{\partial\mu}\Gamma_S(p,p). \end{aligned} \tag{25}$$

Again applying Weinberg's theorem, we find that the entire right-hand side of Eq. (25) vanishes asymptotically as $p^{-1}\times$(powers of $\ln p^2$), and so substituting Eqs. (17) and (18) we get

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma-\alpha\right)J(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} 0. \tag{26}$$

Equation (26) may be immediately integrated to give

$$J(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} f_3(\mu^2/m^2, e_0)(-p^2/m^2)^{(\gamma-\alpha)/2}. \tag{27}$$

Substituting into Eq. (19b) and doing a final integration, we get

$$G(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} K(p^2/m^2)^{(\gamma+1)/2}-f_3(\mu^2/m^2, e_0)(-p^2/m^2)^{(\gamma-\alpha)/2}, \tag{28}$$

with the first term a solution of the *homogeneous* equation

$$\left(m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\gamma\right)mG(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} 0. \tag{29}$$

To determine K, we note that $\gamma\propto e_0^2$ [see Eq. (15b)]. Hence when expanded in powers of e_0^2 the first term in Eq. (28) has, to any finite order in e_0^2, the form

$$K\left(\frac{p^2}{m^2}\right)^{1/2}\times(\text{powers of }\ln p^2), \tag{30}$$

which violates the Weinberg-theorem asymptotic behavior of G stated in Eq. (18). So we conclude that $K=0$, giving

$$G(p^2/m^2, \mu^2/m^2, e_0) \underset{p\to\infty}{\sim} -f_3(\mu^2/m^2, e_0)(-p^2/m^2)^{(\gamma-\alpha)/2}. \tag{31}$$

Defining

$$f_2(\mu^2/m^2, e_0)=\frac{f_3(\mu^2/m^2, e_0)}{f_1(\mu^2/m^2, e_0)},$$

we may combine our results for the asymptotic behavior of $\tilde{S}_F'(p)^{-1}$ into the form

$$\begin{aligned} \tilde{S}_F'(p)^{-1} \underset{p\to\infty}{\sim} & f_1(\mu^2/m^2, e_0)(-p^2/m^2)^{\gamma/2} \\ &\times[\gamma\cdot p-mf_2(\mu^2/m^2, e_0)(-p^2/m^2)^{-\alpha/2}]. \end{aligned} \tag{32}$$

Our final step is to show that γ and α are independent of μ^2/m^2. We do this by writing down the analogs of Eq. (15a) and Eq. (24) obtained by differentiating with respect to the photon mass μ only, with the electron mass m held fixed. These are

$$\begin{aligned} &\left(\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right)\tilde{S}_F'(p)^{-1}=-m\alpha_\mu\tilde{\Gamma}_S(p,p)-\mu^2\tilde{\Gamma}_{S'}(p,p), \\ &\left(\mu\frac{\partial}{\partial\mu}+\gamma_\mu-\alpha_\mu\right)\tilde{\Gamma}_S(p,p)=m\alpha_\mu\tilde{T}_{SS}(p,p)+\mu^2\tilde{T}_{SS'}(p,p), \\ &\gamma_\mu=-Z_2^{-1}\mu\frac{\partial}{\partial\mu}Z_2, \qquad \alpha_\mu=m_0^{-1}\mu\frac{\partial}{\partial\mu}m_0. \end{aligned} \tag{33}$$

Evaluating Eq. (33) asymptotically, substituting the

results of Eqs. (27) and (32), and separating the terms proportional to $\gamma \cdot p$ and m, we find the two equations[11]

$$\left(\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right)f_1+f_1\ln\left(\frac{p^2}{m^2}\right)\mu\frac{\partial}{\partial\mu}(\tfrac{1}{2}\gamma)\underset{p\to\infty}{\approx}0,$$
$$\left(\mu\frac{\partial}{\partial\mu}+\gamma_\mu-\alpha_\mu\right)f_1f_2+f_1f_2\ln\left(\frac{p^2}{m^2}\right)\mu\frac{\partial}{\partial\mu}\tfrac{1}{2}(\gamma-\alpha)\underset{p\to\infty}{\approx}0. \quad (34)$$

These equations can be satisfied only if the logarithmic terms vanish separately, which implies

$$\frac{\partial}{\partial\mu}\gamma=\frac{\partial}{\partial\mu}\alpha=0, \quad (35a)$$

giving the desired result

$$\gamma=\gamma(e_0), \quad \alpha=\alpha(e_0). \quad (35b)$$

The vanishing of the constant terms in Eq. (34) then implies

$$\left(\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right)f_1=0,$$
$$\left(\mu\frac{\partial}{\partial\mu}-\alpha_\mu\right)f_2=0. \quad (36)$$

On substituting Eq. (22) for m_0 and Z_2 into Eq. (33) for γ_μ and α_μ, and then inserting the resulting expressions into Eq. (36), we find the equations

$$\frac{\partial}{\partial\mu}(f_1/C_1)=\frac{\partial}{\partial\mu}(f_2/C_2)=0,$$
$$f_1/C_1=F_1(e_0), \quad f_2/C_2=F_2(e_0). \quad (37)$$

Hence our final result for the asymptotic behavior of the electron propagator is

$$\tilde{S}_F'(p)^{-1}\underset{p\to\infty}{\approx}F_1(e_0)C_1(\mu^2/m^2,e_0)(-p^2/m^2)^{\gamma(e_0)/2}[\gamma\cdot p-mF_2(e_0)C_2(\mu^2/m^2,e_0)(-p^2/m^2)^{-\alpha(e_0)/2}], \quad (38)$$

with C_1 and C_2 the same functions as appear in Eq. (22) for Z_2 and m_0.

The result of Eq. (38) is valid for arbitrary values of the gauge parameter ξ in Eq. (7).[12] Under the change of gauge

$$\left(\frac{\xi q_\mu q_\nu}{q^2}-g_{\mu\nu}\right)\frac{1}{q^2-\mu^2+i\epsilon}\,\frac{-\Lambda^2}{q^2-\Lambda^2+i\epsilon}$$
$$\to\left(\frac{\xi' q_\mu q_\nu}{q^2}-g_{\mu\nu}\right)\frac{1}{q^2-\mu^2+i\epsilon}\,\frac{-\Lambda^2}{q^2-\Lambda^2+i\epsilon} \quad (39)$$

it is easily shown[13] that the unrenormalized position-space electron propagator transforms in the simple fashion

$$S_F'(x)\to\exp\{(\xi'-\xi)[\lambda(x)-\lambda(0)]\}S_F'(x),$$
$$\lambda(x)=\frac{ie_0^2}{(2\pi)^4}\int\frac{d^4q}{q^2}\,\frac{1}{q^2-\mu^2}\,\frac{-\Lambda^2}{q^2-\Lambda^2}e^{-iq\cdot x}. \quad (40)$$

By studying the large-x and small-x behavior of Eq. (40), we can find the behavior under gauge transformation of the quantities F_1, C_1, F_2, C_2, α, and γ appearing in Eqs. (22) and (38). Suppressing the dependence on μ^2/m^2 and e_0, and letting primed quantities denote those computed with gauge parameter ξ' and unprimed quantities those computed with gauge parameter ξ, we find

$$\gamma'-\gamma=(\alpha_0/2\pi)(\xi'-\xi),$$
$$\alpha'-\alpha=0,$$
$$\frac{C_1'}{C_1}=\left(\frac{m^2}{\mu^2}\right)^{(\alpha_0/4\pi)(\xi'-\xi)}, \quad \frac{C_2'}{C_2}=1,$$
$$\frac{Z_2'}{Z_2}=\left(\frac{\Lambda^2}{\mu^2}\right)^{(\alpha_0/4\pi)(\xi'-\xi)}, \quad \frac{m_0'}{m_0}=1; \quad (41a)$$

$$\frac{F_1'}{F_1}=\frac{1-\frac{1}{2}\gamma'}{1-\frac{1}{2}\gamma}\,\frac{\Gamma(1+\frac{1}{2}\gamma)\Gamma(1-\frac{1}{2}\gamma')}{\Gamma(1-\frac{1}{2}\gamma)\Gamma(1+\frac{1}{2}\gamma')}$$
$$\times\exp[(-\alpha_0/4\pi)(2\gamma_E-1)(\xi'-\xi)],$$
$$\frac{F_2'}{F_2}\frac{F_1}{F_1'}=\frac{\Gamma(1-\frac{1}{2}\gamma-\frac{1}{2}\alpha)\Gamma(1+\frac{1}{2}\gamma'+\frac{1}{2}\alpha')}{\Gamma(1+\frac{1}{2}\gamma+\frac{1}{2}\alpha)\Gamma(1-\frac{1}{2}\gamma'-\frac{1}{2}\alpha')}$$
$$\times\exp[(\alpha_0/4\pi)(2\gamma_E-1)(\xi'-\xi)], \quad (41b)$$

with $\gamma_E=0.577\,21\ldots=$ Euler's constant and with Γ the usual Γ function. The derivation leading to Eq. (41) is given in Appendix B.

According to Eqs. (22) and (41a), if we choose ξ' to satisfy $(\alpha_0/2\pi)(\xi'-\xi)+\gamma(e_0,\xi)=0$, then we have $\gamma'\equiv\gamma(e_0,\xi')=0$ and the wave-function renormalization Z_2' becomes finite as $\Lambda\to\infty$. This choice of gauge (the Landau gauge) is the one used by Baker and Johnson in their work. In the Landau gauge, Eq. (38) becomes

$$\tilde{S}_F'(p)^{-1}_{\text{Landau gauge}}\underset{p\to\infty}{\approx}C[\gamma\cdot p+am(m^2/-p^2)^\epsilon],$$
$$C=F_1'C_1', \quad a=-F_2'C_2', \quad \epsilon=\tfrac{1}{2}\alpha(e_0), \quad (42)$$

in agreement with the Baker-Johnson result stated in Eq. (2). We note also that the gauge-independent quantity $m_0=m-\Sigma|_{\gamma\cdot p=m}$ is finite as $\mu\to0$ [only $Z_2=1-\partial\Sigma/\partial(\gamma\cdot p)|_{\gamma\cdot p=m}$ is infrared-divergent]; hence the function $C_2(\mu^2/m^2,e_0)$ appearing in Eqs. (22) and (38) has a finite limit as $\mu\to0$. Let us give two simple second-order calculations which illustrate Eqs. (42) and (38). First, we calculate ϵ in Eq. (42) by noting that to second order

$$1+\alpha(e_0)=\frac{m}{m_0}\left(\frac{\partial}{\partial m}+\frac{\mu}{m}\frac{\partial}{\partial\mu}\right)m_0$$
$$\approx 1+\frac{\Sigma(m)}{m}-\left(\frac{\partial}{\partial m}+\frac{\mu}{m}\frac{\partial}{\partial\mu}\right)\Sigma(m),$$
$$\Sigma(m)=\Sigma|_{\gamma\cdot p=m_0=m}\,. \tag{43}$$

Explicit calculation in an arbitrary covariant gauge shows that

$$\Sigma(m)=m[(3\alpha_0/4\pi)\ln(\Lambda^2/m^2)+\text{function of }(\mu^2/m^2)], \tag{44}$$

which on substitution into Eq. (43) gives

$$\alpha(e_0)=3\alpha_0/2\pi\,,\quad \epsilon=\tfrac{3}{2}\alpha_0/2\pi\,, \tag{45}$$

in agreement with the second-order term in Eq. (3). As our second illustration we show that, to second order, m_0, Z_2, and the full renormalized electron propagator $\tilde{S}_F'(p)^{-1}$ in the Feynman gauge do satisfy Eqs. (22) and (38). A straightforward calculation gives

$$\tilde{S}_F'(p)^{-1}=Z_2(\gamma\cdot p-m_0)-\frac{\alpha_0}{2\pi}\int_0^1 dz[(1-z)\gamma\cdot p-2m]\ln\left(\frac{zm^2+(1-z)\mu^2-z(1-z)p^2}{(1-z)\Lambda^2}\right)$$
$$=\left[1+\frac{\alpha_0}{2\pi}\int_0^1 dz\frac{z(1-z^2)2m^2}{z^2m^2+(1-z)\mu^2}-\frac{\alpha_0}{2\pi}\int_0^1 dz(1-z)\ln\left(\frac{zm^2+(1-z)\mu^2-z(1-z)p^2}{z^2m^2+(1-z)\mu^2}\right)\right]$$
$$\times\left\{\gamma\cdot p-m\left[1-\frac{\alpha_0}{2\pi}\int_0^1 dz(1+z)\ln\left(\frac{zm^2+(1-z)\mu^2-z(1-z)p^2}{z^2m^2+(1-z)\mu^2}\right)\right]\right\}, \tag{46}$$

from which we can identify the quantities appearing in Eqs. (22) and (38),

$$\gamma(e_0)=-\alpha_0/2\pi\,,\quad \alpha(e_0)=3\alpha_0/2\pi\,,\quad F_1(e_0)=1+3\alpha_0/8\pi\,,$$
$$C_1(\mu^2/m^2,e_0)=1+\frac{\alpha_0}{2\pi}\int_0^1 dz\left[(1-z)\ln\left(\frac{z^2m^2+(1-z)\mu^2}{(1-z)m^2}\right)+\frac{z(1-z^2)2m^2}{z^2m^2+(1-z)\mu^2}\right],$$
$$F_2(e_0)=1+5\alpha_0/8\pi\,,$$
$$C_2(\mu^2/m^2,e_0)=1+\frac{\alpha_0}{2\pi}\int_0^1 dz(1+z)\ln\left(\frac{z^2m^2+(1-z)\mu^2}{(1-z)m^2}\right). \tag{47}$$

As $\mu\to 0$, we see that $C_2(\mu^2/m^2,e_0)$ approaches a finite limit, as expected.

This completes our treatment of the electron propagator. Let us next turn briefly to the asymptotic behavior of the photon propagator.[14] Because renormalization of the photon propagator is subtractive, i.e.,

$$\tilde{\pi}(q^2)=\pi(q^2)-\pi(0)\,, \tag{48}$$

the renormalized photon self-energy $\tilde{\pi}$ involves, even for asymptotically large q^2, the nonasymptotic piece $\pi(0)$. As a result, the asymptotic behavior of the *renormalized* photon propagator cannot be calculated by replacing all internal photon propagators by their asymptotic form, Eq. (1). On the other hand, the asymptotic *unrenormalized* photon self-energy $\pi(q^2)$ *does* involve only the asymptotic behavior of the internal photon propagator, and thus can be calculated in our model in which internal photon lines are replaced by Eq. (1). To proceed, we write down the two Callan-Symanzik scaling equations obtained by differentiating π with respect to m and with respect to μ. These are

$$m\frac{\partial}{\partial m}\pi(q^2/\Lambda^2,q^2/\mu^2,q^2/m^2,e_0)$$
$$=m(1+\alpha-\alpha_\mu)\pi_S(q^2/\Lambda^2,q^2/\mu^2,q^2/m^2,e_0)\,,$$
$$\mu\frac{\partial}{\partial\mu}\pi(q^2/\Lambda^2,q^2/\mu^2,q^2/m^2,e_0)$$
$$=m\alpha_\mu\pi_S(q^2/\Lambda^2,q^2/\mu^2,q^2/m^2,e_0)+\mu^2\pi_{S'}(q^2/\Lambda^2,q^2/\mu^2,q^2/m^2,e_0)\,, \tag{49}$$

with α and α_μ given by Eqs. (15b) and (33) above and with the photon-photon-scalar three-point functions π_S and $\pi_{S'}$ given by

$$m\pi_S=m_0\frac{\partial}{\partial m_0}\pi\,,$$
$$\mu^2\pi_{S'}=\mu\frac{\partial}{\partial\mu}\pi\,. \tag{50}$$

Let us now consider the asymptotic limit in which q and Λ both become large. Application of Weinberg's theorem[15] shows that π_S and $\pi_{S'}$ vanish as $q^{-1}\times$(powers of $\ln q^2$) and $q^{-2}\times$(powers of $\ln q^2$), respectively, so that the right-hand sides of the scaling equations can be neglected, giving

$$m\frac{\partial}{\partial m}\pi(q^2/\Lambda^2, q^2/\mu^2, q^2/m^2, e_0) \underset{q,\Lambda\to\infty}{\sim} 0\,,$$
$$\mu\frac{\partial}{\partial \mu}\pi(q^2/\Lambda^2, q^2/\mu^2, q^2/m^2, e_0) \underset{q,\Lambda\to\infty}{\sim} 0\,. \tag{51}$$

This tells us that the asymptotic unrenormalized photon self-energy has no dependence on m and μ, that is,

$$\pi(q^2/\Lambda^2, q^2/\mu^2, q^2/m^2, e_0) \underset{q,\Lambda\to\infty}{\sim} \pi(q^2/\Lambda^2, e_0)\,. \tag{52}$$

Furthermore, since to any finite order of perturbation theory the dependence of π on Λ^2 can only be through powers of $\ln\Lambda^2$, then $\pi(q^2/\Lambda^2, e_0)$ must have the form

$$\pi(q^2/\Lambda^2, e_0) = \sum_{n=0}^{\infty} B_n(e_0)[\ln(-q^2/\Lambda^2)]^n\,. \tag{53}$$

We now invoke the fact that since π is gauge-independent we are free to choose the gauge which makes Z_2 finite, and since π contains no internal photon self-energy parts, the subintegrations of π which do not involve all lines in the graph are finite. But the single subintegration which does involve all lines is made finite by a single differencing of π,

$$\pi(q^2/\Lambda^2, e_0) - \pi(q_1{}^2/\Lambda^2, e_0) = \text{cutoff-independent}\,, \tag{54}$$

which tells us that only the $n=0$ and $n=1$ terms can be present in Eq. (53). Thus,

$$\pi(q^2/\Lambda^2, q^2/\mu^2, q^2/m^2, e_0) \underset{q,\Lambda\to\infty}{\sim} B_0(e_0) + B_1(e_0)\ln(-q^2/\Lambda^2)\,, \tag{55}$$

in agreement with the result of JBW stated in Eq. (5), with $f(\alpha_0) = B_1(e_0)$ the function which determines the eigenvalue condition.

We wish to acknowledge the hospitality of the Aspen Center for Physics, where this work was done.

APPENDIX A

We give here a proof that $m\tilde{\Gamma}_S(p,p) \equiv m_0 Z_2 \Gamma_S(p,p)$ is finite (cutoff-independent as $\Lambda\to\infty$) to all orders of perturbation theory. Let us define $\Gamma^5(p,p)$ to be the zero-momentum-transfer vertex of the pseudoscalar electron current $j^5 = \bar{\psi}\gamma^5\psi$. We have previously shown,[16] by using the axial-vector–vertex Ward identity, that $m\tilde{\Gamma}^5(p,p) \equiv m_0 Z_2 \Gamma^5(p,p)$ is finite to all orders of perturbation theory.[17] Let us define

$$\Delta(p,p) = \tilde{\Gamma}_S(p,p)\gamma^5 - \tilde{\Gamma}^5(p,p)\,. \tag{A1}$$

To zeroth order in perturbation theory, $\Delta(p,p)=0$ is finite. Let us now make the inductive hypothesis that, to order $n-2$ in perturbation theory, (i) $\Delta(p,p)$ is finite and (ii) as $p\to\infty$, $\Delta(p,p)\sim p^{-1}\times$(powers of $\ln p^2$). To prove that these hypotheses are satisfied in order n as well, we follow very closely the procedure used in Chap. 19 of Ref. 3 to prove that the usual renormalizations of electrodynamics make the vector vertex finite. We begin by observing that $m\tilde{\Gamma}_S$ and $m\tilde{\Gamma}^5$ satisfy the integral equations (see Fig. 4)

$$m\tilde{\Gamma}_S = m_0 Z_2 - \int m\tilde{\Gamma}_S \tilde{S}_F'\tilde{S}_F'\tilde{K}\,,$$
$$m\tilde{\Gamma}^5 = m_0 Z_2\gamma^5 - \int m\tilde{\Gamma}^5 \tilde{S}_F'\tilde{S}_F'\tilde{K}\,, \tag{A2}$$

with $\tilde{K}$ the connected, renormalized electron-positron scattering kernel, obtained by excluding the class of graphs shown in Fig. 5. Substituting Eq. (A2) into Eq. (A1), we find that the inhomogeneous terms cancel, giving the following expression for Δ:

$$\Delta = \int \tilde{\Gamma}^5\tilde{S}_F'\tilde{S}_F'\tilde{K} - \int \tilde{\Gamma}_S\tilde{S}_F'\tilde{S}_F'\tilde{K}\gamma^5\,. \tag{A3}$$

Since the perturbation expansion of $\tilde{K}$ begins in second order, to calculate Δ to order n we need only insert $\tilde{\Gamma}^5$ and $\tilde{\Gamma}_S$ to order $n-2$ on the right-hand side of Eq. (A3). But these are known to be finite by the inductive hypothesis, so the individual factors appearing on the right-hand side of Eq. (A3) are finite. According to Weinberg's theorem, to see whether Δ is finite to order n, and to determine its large-p asymptotic behavior, we must determine the naive degree of divergence D of each subintegration contributing to the right-hand side of Eq. (A3). As is shown on pp. 330–334 of Ref. 3, all subintegrations have $D\leq -1$, except possibly those involving both electron propagators $\tilde{S}_F'$ and all lines in the kernel $\tilde{K}$. These are of two basic types, according to whether the electron-positron lines emerging from $\tilde{\Gamma}_S$ and $\tilde{\Gamma}^5$ *do* [Fig. 6(a)] or *do not* [Fig. 6(b)] connect directly with the external electron-positron lines entering the kernel $\tilde{K}$. Clearly, the diagrams shown in Fig. 6(b) involve a closed electron loop with a single scalar or pseudoscalar vertex. Charge-conjugation invariance implies that such a loop can have only an even number of photon vertices and an odd number of electron propagators; the fact that the trace of an odd number of γ matrices vanishes then implies that such loops are proportional to the electron mass

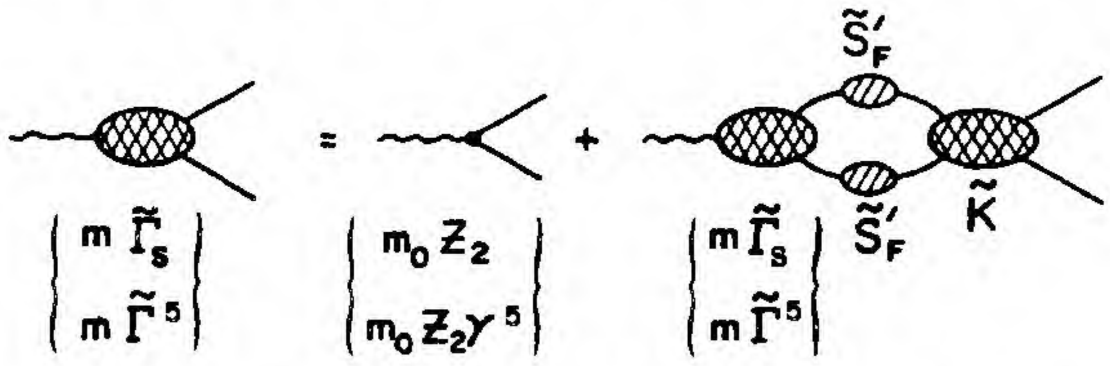


FIG. 4. Integral equations satisfied by the scalar and pseudoscalar vertex parts.

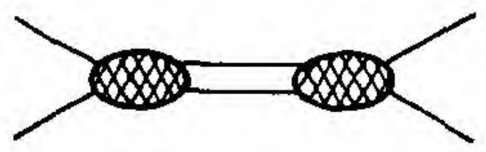


FIG. 5. Class of graphs excluded from the kernel $\tilde{K}$.

m. So the diagrams shown in Fig. 6(b) are all proportional to m, which improves the convergence by one power of momentum and gives them $D \leq -1$. The contribution to Eq. (A3) of the diagrams shown in Fig. 6(a) can be written symbolically as

$$\begin{aligned}\Delta^{6(a)} &= \int (\tilde{\Gamma}^5 - \tilde{\Gamma}_S\gamma^5)\tilde{S}_F'\tilde{S}_F'\tilde{K} \\ &\quad - \int \tilde{\Gamma}_S[\tilde{S}_F'\tilde{S}_F'\tilde{K}\gamma^5 - \gamma^5\tilde{S}_F'\tilde{S}_F'\tilde{K}] \\ &= \int \Delta\tilde{S}_F'\tilde{S}_F'\tilde{K} - \int \tilde{\Gamma}_S[\tilde{S}_F'\tilde{S}_F'\tilde{K}\gamma^5 - \gamma^5\tilde{S}_F'\tilde{S}_F'\tilde{K}]\,. \end{aligned} \tag{A4}$$

The first term in Eq. (A4) has $D \leq -1$ because (to order $n-2$) Δ satisfies assumption (ii) of the inductive hypothesis. The second term in Eq. (A4) is the residue obtained when the matrix γ^5 is commuted from its original position on the far right of Fig. 6(a), through the string of electron propagators and photon vertices, to a position immediately to the right of the vertex $\tilde{\Gamma}_S$. The square bracket in this term is easily seen to be proportional to the electron mass m, giving the second term an extra power of convergence with the result that it, too, has $D \leq -1$. This completes the demonstration that, to order n, all subintegrations

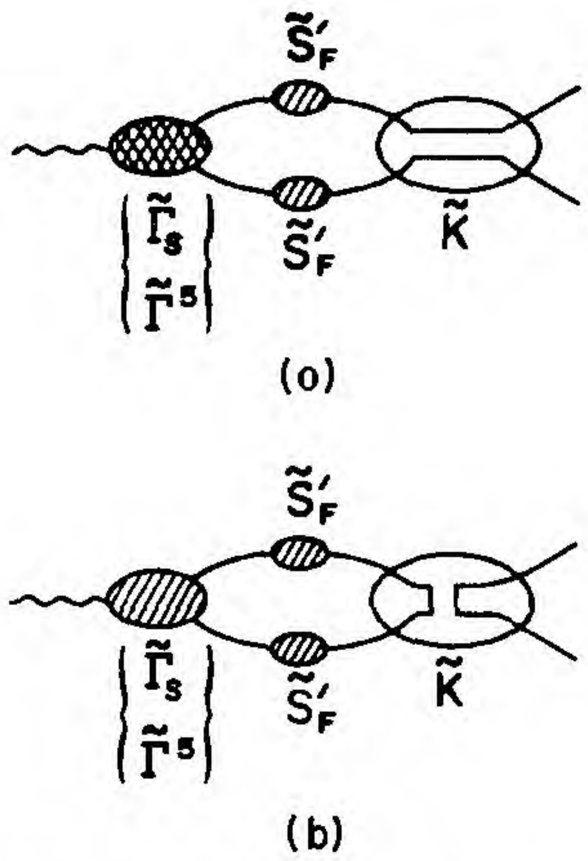


FIG. 6. (a) Diagrams in which the electron-positron lines emerging from $\tilde{\Gamma}_S$ and $\tilde{\Gamma}^5$ connect directly with the external electron-positron lines entering the kernel $\tilde{K}$. Internal photon lines in $\tilde{K}$ are not shown. (b) Diagrams in which the electron-positron lines emerging from $\tilde{\Gamma}_S$ and $\tilde{\Gamma}^5$ *do not* connect directly with the external electron-positron lines entering the kernel $\tilde{K}$. Again, internal photon lines in $\tilde{K}$ are not shown.

contributing to the right-hand side of Eq. (A3) have $D \leq -1$. By Weinberg's theorem, this implies that, to order n, Δ has the properties (i) and (ii) stated above, thereby completing the induction.

We note that in making the proof we have not assumed the omission of internal photon self-energy parts. Our result, that Δ is finite, is *a fortiori* still valid when this simplification is made.

APPENDIX B

We derive here the results quoted in Eq. (41) of the text, giving the behavior under gauge transformation of the quantities F_1, C_1, F_2, C_2, α, and γ. Our starting point is Eq. (40), which we repeat for convenience:

$$\begin{aligned} S_F'(x,\xi') &= \exp\{(\xi'-\xi)[\lambda(x)-\lambda(0)]\}S_F'(x,\xi)\,, \\ \lambda(x) &= \frac{ie_0^2}{(2\pi)^4}\int\frac{d^4q}{q^2}\,\frac{1}{q^2-\mu^2}\,\frac{-\Lambda^2}{q^2-\Lambda^2}\,e^{-iq\cdot x}\,. \end{aligned} \tag{B1}$$

Before proceeding with the derivation, we give some useful properties of $\lambda(x)$. Letting x be spacelike, and using the symmetrical integration formula

$$\int d\Omega_q e^{-iq\cdot x} = \frac{4\pi^2}{(-q^2)^{1/2}(-x^2)^{1/2}}\,J_1((-q^2)^{1/2}(-x^2)^{1/2})\,, \tag{B2}$$

(J_1 = Bessel function of order unity), we find the following representation for $\lambda(x)$:

$$\lambda(x) = \frac{-\alpha_0}{\pi}\int_0^\infty \frac{d\rho}{\rho^2 - x^2\mu^2}\left(\frac{-x^2\Lambda^2}{\rho^2-x^2\Lambda^2}\right)J_1(\rho)\,. \tag{B3}$$

From Eqs. (B1) and (B3) we learn that

$$\begin{aligned} &\lambda(x) \underset{x\to\infty}{\approx} 0\,, \\ &\lambda(x) \underset{x\to 0}{\approx} \frac{-\alpha_0}{4\pi}\ln(\Lambda^2/\mu^2) + O(\mu^2/\Lambda^2) + O(x^2\ln x^2)\,; \\ &\tilde{\lambda}(x) \equiv \lim_{\Lambda\to\infty}\lambda(x) = \frac{-\alpha_0}{\pi}\int_0^\infty\frac{d\rho}{\rho^2-x^2\mu^2}J_1(\rho)\,, \\ &\tilde{\lambda}(x) \underset{x\to 0}{\approx} \frac{\alpha_0}{4\pi}[\ln(-\tfrac{1}{4}x^2\mu^2) + 2\gamma_E - 1] + O(x^2\ln x^2)\,, \end{aligned} \tag{B4}$$

γ_E = Euler's constant.

We begin by deriving the results of Eq. (41a), giving the gauge-transformation behavior of the renormalization constants m_0 and Z_2. Introducing the wave-function renormalization $Z_2(\xi)$ and the renormalized electron propagator $\tilde{S}_F'(x,\xi)$,

$$\tilde{S}_F'(x,\xi) = Z_2(\xi)^{-1}S_F'(x,\xi)\,, \tag{B5}$$

we can rewrite Eq. (B1) in the form

$$\begin{aligned}\tilde{S}_F'(x,\xi') &= Z_2(\xi)Z_2(\xi')^{-1} \\ &\quad\times\exp\{(\xi'-\xi)[\lambda(x)-\lambda(0)]\}\tilde{S}_F'(x,\xi)\,. \end{aligned} \tag{B6}$$

Let us consider the limit of Eq. (B6) as $x\to\infty$.

Because we have supplied an infrared cutoff μ, the renormalized electron propagator approaches in this limit the free electron propagator for physical mass m,

$$\tilde{S}_F'(x,\xi) \underset{x\to\infty}{\longrightarrow} S_F^0(x) = \int \frac{d^4p}{(2\pi)^4} \frac{e^{-ip\cdot x}}{\gamma\cdot p - m}. \tag{B7}$$

Since the right-hand side of Eq. (B7) is independent of gauge, we get, using the results of Eq. (B4),

$$\begin{aligned}\frac{Z_2'}{Z_2} \equiv \frac{Z_2(\xi')}{Z_2(\xi)} &= \exp[-(\xi'-\xi)\lambda(0)] \\ &= \left(\frac{\Lambda^2}{\mu^2}\right)^{(\alpha_0/4\pi)(\xi'-\xi)}.\end{aligned} \tag{B8}$$

Comparing Eq. (B8) with Eq. (22) in the text, we learn that

$$\begin{aligned}\gamma' - \gamma &= (\alpha_0/2\pi)(\xi'-\xi), \\ \frac{C_1'}{C_1} &= \left(\frac{m^2}{\mu^2}\right)^{(\alpha_0/4\pi)(\xi'-\xi)}.\end{aligned} \tag{B9}$$

To get the gauge transformation properties of the bare mass m_0, we consider the small-x limit of the unrenormalized equation, Eq. (B1). The small-x behavior of the unrenormalized position-space propropagator is determined by the large-p behavior of the unrenormalized momentum-space propagator by the Fourier-transform relation

$$S_F'(x,\xi) = \int \frac{d^4p}{(2\pi)^4} \frac{e^{-ip\cdot x}}{\gamma\cdot p - m_0(\xi) - \Sigma(p,\xi)}. \tag{B10}$$

But as $p\to\infty$ for fixed cutoff Λ, we have

$$\Sigma(p,\xi) \underset{p\gg\Lambda}{\sim} p^{-1}\times(\text{powers of } \ln p^2), \tag{B11}$$

so we get from Eq. (B10)

$$\begin{aligned}S_F'(x,\xi) &= \int \frac{d^4p}{(2\pi)^4} \frac{e^{-ip\cdot x}}{\gamma\cdot p - m_0(\xi) + O(p^{-1}\times(\text{powers of } \ln p^2))} \\ &= \int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot x}\left[\frac{\gamma\cdot p}{p^2} + \frac{m_0(\xi)}{p^2} + O(p^{-3}\times(\text{powers of } \ln p^2))\right] \\ &= \frac{1}{2\pi^2}\frac{\gamma\cdot x}{(x^2)^2} + \frac{i}{4\pi^2}\frac{m_0(\xi)}{x^2} + O(x^{-1}\times(\text{powers of } \ln x^2)).\end{aligned} \tag{B12}$$

Substituting Eq. (B12) into the small-x limit of Eq. (B1), and using Eq. (B4) for a small-x estimate of $\lambda(x) - \lambda(0)$, we get

$$\frac{m_0'}{m_0} \equiv \frac{m_0(\xi')}{m_0(\xi)} = 1. \tag{B13}$$

Comparing with Eq. (22) in the text, we find

$$\alpha' = \alpha, \quad \frac{C_2'}{C_2} = 1. \tag{B14}$$

Next, we derive the results of Eq. (41b), giving the gauge-transformation behavior of the functions F_1 and F_2 appearing in the asymptotic form of the renormalized propagator. To proceed, we need the renormalized version of Eq. (B1), obtained by eliminating $Z_2(\xi)Z_2(\xi')^{-1}$ from Eq. (B6) by use of Eq. (B8) and then dropping terms in $\lambda(x)$ which, for fixed x, vanish as $\Lambda\to\infty$. This gives

$$\tilde{S}_F'(x,\xi') = \exp[(\xi'-\xi)\tilde{\lambda}(x)]\,\tilde{S}_F'(x,\xi), \tag{B15}$$

with $\tilde{\lambda}(x)$ given in Eq. (B4). We now take the small-x limit of Eq. (B15), using Eq. (B4) for the small-x behavior of $\tilde{\lambda}(x)$ and extracting the small-x behavior of $\tilde{S}_F'(x,\xi)$ from the large-p asymptotic form of $\tilde{S}_F'^{-1}(p,\xi)$ given in Eq. (38),

$$\tilde{S}_F'(x,\xi) = \int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot x}\left[\frac{(-p^2/m^2)^{-\gamma/2}}{F_1C_1}\frac{\gamma\cdot p + mF_2C_2(-p^2/m^2)^{-\alpha/2}}{p^2} + O(p^{-3}\times(\text{powers of } \ln p^2))\right]. \tag{B16}$$

We can evaluate the integrals appearing in Eq. (B15) by using Eq. (B2) and the formula[18]

$$\int_0^\infty dx\, x^{-\gamma}J_1(x) = 2^{-\gamma}\frac{\Gamma(1-\frac{1}{2}\gamma)}{\Gamma(1+\frac{1}{2}\gamma)}. \tag{B17}$$

Substituting the result into Eq. (B4) and equating the coefficients on left and right of the two most singular terms as $x\to 0$, we get the gauge-transformation formulas quoted in Eq. (41b) of the text.

*Research supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U. S. Air Force under Contract No. F44620-71-C-0044.

[1]K. Johnson, M. Baker, and R. Willey, Phys. Rev. 136, B1111 (1964); 163, 1699 (1967); M. Baker and K. Johnson, *ibid.* 183, 1292 (1969); Phys. Rev. D 3, 2516, 2541 (1971).

[2]M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954).

[3]Throughout this paper we use the notation and metric conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), pp. 377-390.

[4]S. Weinberg, Phys. Rev. 118, 838 (1960).

[5]This result for the asymptotic *renormalized* electron propagator follows from the facts that (i) it is true for the unrenormalized electron propagator, and (ii) renormalization of the electron propagator is *multiplicative;* $\tilde{S}'_F(p)^{-1}=Z_2 S'_F(p)^{-1}$. As pointed out below, because renormalization of the photon self-energy part is *subtractive*, an analogous result cannot be obtained for the *renormalized* photon self-energy part.

[6]C. G. Callan, Phys. Rev. D 2, 1541 (1970); K. Symanzik, Commun. Math. Phys. 18, 227 (1970).

[7]This point has been emphasized by S. Coleman and R. Jackiw, Ann. Phys. (N.Y.) (to be published), and by A. H. Mueller and T. L. Trueman, Phys. Rev. D 4, 1635 (1971).

[8]S. Coleman (unpublished) and S. Coleman and R. Jackiw, Ref. 7.

[9]When photon self-energy parts are included, the left-hand side of Eq. (15a) acquires an additional term proportional to $(\partial/\partial e)\tilde{S}'_F(p)^{-1}$, where e is the renormalized charge.

[10]The finiteness of γ and α also follows directly from the inductive derivation of the scaling equations given by Callan in Ref. 6.

[11]In obtaining the two relations in Eq. (34), we have divided by $(-p^2/m^2)^{\gamma/2}$ and $(-p^2/m^2)^{(\gamma-\alpha)/2}$, respectively. Since the exponents $\gamma/2$ and $(\gamma-\alpha)/2$ are both proportional to $e_0{}^2$, the factors which we have divided out are simply powers of $\ln(-p^2/m^2)$ to any finite order of perturbation theory. Hence the right-hand sides of the relations in Eq. (34) remain of order $p^{-1}\times$(powers of $\ln p^2$), and thus can be safely neglected relative to the left-hand sides.

[12]It is actually true for more general gauges than the simple covariant gauge used in Eq. (7).

[13]L. D. Landau and I. M. Khalatnikov, Zh. Eksperim. i Teor. Fiz. 29, 89 (1955) [Soviet Phys. JETP 2, 69 (1956)]; K. Johnson and B. Zumino, Phys. Rev. Letters 3, 351 (1959).

[14]Similar work has been done by N. Christ (unpublished).

[15]Recall that π_S and $\pi_{S'}$ are obtained by removing a gauge-invariant factor $q^2 g_{\mu\nu}-q_\mu q_\nu$ from the full Feynman photon-photon-scalar three-point amplitudes. Thus, π_S and $\pi_{S'}$ vanish asymptotically two powers faster than the corresponding full three-point amplitudes.

[16]S. L. Adler and W. A. Bardeen, Phys. Rev. 182, 1517 (1969).

[17]We are of course assuming the fact that the usual renormalizations of electrodynamics make finite all n-point functions *not* involving the scalar or pseudoscalar vertex.

[18]The Bateman Manuscript Project, *Tables of Integral Transforms*, edited by A. Erdélyi (McGraw-Hill, New York, 1954), Vol. 2, p. 22, Eq. (7).

PHYSICAL REVIEW D VOLUME 5, NUMBER 12 15 JUNE 1972

Short-Distance Behavior of Quantum Electrodynamics and an Eigenvalue Condition for α

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 31 January 1972)

We review and extend earlier work dealing with the short-distance behavior of quantum electrodynamics. We show that if the renormalized photon propagator is asymptotically finite, then in the limit of zero fermion mass all of the single-fermion-loop $2n$-point functions, regarded as functions of the coupling constant, must have a common infinite-order zero. In the usual class of asymptotically finite solutions introduced by Gell-Mann and Low, the asymptotic coupling α_0 is fixed to be this infinite-order zero and the physical coupling $\alpha < \alpha_0$ is a free parameter. We show that if the single-fermion-loop diagrams actually possess the required infinite-order zero, there is a unique, additional solution in which the physical coupling α is fixed to be the infinite-order zero. We conjecture that this is the solution chosen by nature. According to our conjecture, the fine-structure constant is determined by the eigenvalue condition $F^{[1]}(\alpha) = 0$, where $F^{[1]}$ is a function related to the single-fermion-loop vacuum-polarization diagrams. The eigenvalue condition is independent of the number of fundamental fermion species which are assumed to be present.

I. INTRODUCTION AND SUMMARY

The fundamental constant regulating all microscopic electronic phenomena, from atomic physics to quantum electrodynamics, is the fine-structure constant α. Experimentally, the current value[1] $\alpha = 1/(137.03602 \pm 0.00021)$ is one of the best determined numbers in physics. Theoretically, the reason why nature selects this particular numerical value has remained a mystery, and has provoked much interesting speculation. The speculations may be divided roughly into three general types: (a) those in which α is cosmologically determined, either as a cosmological boundary condition (which makes α undeterminable) or as a function of time-varying cosmological parameters (which makes α a function of time)[2]; (b) theories in which α is a constant which is determined microscopically through the interplay of the electromagnetic interaction with interactions of other types, either strong, weak, or gravitational.[3] Since these interactions are currently even less well understood than is the electromagnetic interaction, such theories seem at present to offer little promise of an actual computation of α; (c) finally, theories in which α is microscopically determined through properties of the electromagnetic interaction alone, considered in isolation from other interactions. It is this restricted class of theories to which we will address ourselves in the present paper.

The idea that α may be determined electromagnetically is an old one. In the early days of renormalization theory there were hopes that α could be fixed by requiring the logarithmic divergences appearing in higher orders of perturbation theory to cancel or "compensate" the second-order divergence in the photon wave-function renormalization Z_3,[4] so that the renormalized photon propagator would be asymptotically finite. These hopes received a setback, however, when Jost and Luttinger[5] calculated the order-α^2 logarithmically divergent contribution to Z_3 and found that it has the same sign as the order-α divergence. Of course, it was obvious that the question could not be settled by calculations to any finite order of perturbation theory. A systematic nonperturbative attack on the problem was made by Gell-Mann and Low[6] in their classic 1954 paper on renormalization-group methods. They showed that there is

indeed an eigenvalue condition imposed by requiring that the renormalized photon propagator behave as

$$\alpha d_c(-q^2/m^2, \alpha) = \alpha_0 + h(-q^2/m^2, \alpha), \quad (1)$$

with α_0 finite and with h vanishing as $-q^2/m^2 \to \infty$. However, the condition takes the form $\psi(\alpha_0)=0$, and determines the *asymptotic coupling* α_0 rather than the physical coupling α. Their analysis leaves α a free parameter of the theory, restricted only by the condition $\alpha < \alpha_0$ coming from spectral-function positivity. This essential conclusion was retained in the subsequent important work of Johnson, Baker, and Willey,[7] who showed that if the eigenvalue condition is satisfied *all* the renormalization constants of electrodynamics (m_0 and Z_2 as well as Z_3) can be finite, and who applied a simple argument based on the Federbush-Johnson theorem[8] to obtain a greatly simplified form of the eigenvalue equation for α_0. Thus, the prevailing view since 1954 has been that it is not possible to determine α within a purely electrodynamical context.

Our aim in the present paper is to give a reexamination and extension of the work of Gell-Mann and Low and of Johnson, Baker, and Willey, which, we believe, reopens the possibility of an electrodynamic determination of α. We continue to work within the same basic framework as these previous authors in that we assume, as they do, that asymptotically vanishing terms encountered in each order of perturbation theory do not sum to give an asymptotically dominant result. Our basic observation is that the work of Johnson *et al.*, assumes that α_0 is both a simple zero, and a point of regularity, of the Gell-Mann–Low function $\psi(y)$. In actual fact, we find that an extension of the argument given by Baker and Johnson to obtain their simplified eigenvalue condition indicates that neither of these assumptions is correct. We show that if ψ has a zero at all it must be a zero of infinite order – i.e., an essential singularity. This infinite-order zero in the coupling constant must also appear in all of the single-fermion-loop $2n$-point functions calculated in electrodynamics with zero fermion mass. The presence of an essential singularity has the important consequence that different orders of summing perturbation theory can lead to *inequivalent* forms of the eigenvalue condition. One natural method of summing perturbation theory is to proceed "vacuum-polarization-insertion-wise". One first sums all internal-photon self-energy parts, and then inserts the resulting full photon propagators in the vacuum-polarization skeleton graphs. This order of summation is the one used by Johnson, Baker, and Willey, and leads naturally to the class of asymptotically finite solutions introduced by Gell-Mann and Low, in which α_0 is fixed to be the infinite-order zero and $\alpha < \alpha_0$ is undetermined. An alternative summation method is to proceed "loopwise": One first sums all single-fermion-loop vacuum-polarization graphs, then one sums all two-fermion-loop vacuum-polarization graphs, and so forth. If we assume that the single-fermion-loop $2n$-point functions do actually have the infinite-order zero in the coupling constant as described above, then by using loopwise summation we show that there is a unique additional asymptotically finite solution, in which the physical coupling α is fixed to be the infinite-order zero. We conjecture that this is the solution actually chosen by nature. According to our conjecture, *the fine-structure constant α may be computed as follows. Let $F^{[1]}(y)$ be the coefficient of the logarithmically divergent part of the sum of single-fermion-loop vacuum-polarization diagrams illustrated in Fig. 1. We conjecture that $F^{[1]}(y)$ is analytic in an interval extending from $y=0$ to $y=\alpha$, where it has an infinite-order zero as y approaches α from below along the real axis.* If the function $F^{[1]}(y)$ has no infinite-order zero, then the renormalized photon propagator cannot be asymptotically finite. Our conjecture has the obvious virtue that it stands or falls according to the outcome of the mathematical problem of calculating the function $F^{[1]}(y)$. This problem will be well posed in perturbation theory if $F^{[1]}(y)$ is a function of the class which is uniquely defined by the coefficients of its formal power-series development in y.[9]

The paper is organized as follows. In Sec. II we give a review of previous work on the short-distance behavior of electrodynamics. We derive the Gell-Mann–Low equation for the asymptotic behavior of the photon propagator, discuss its properties, and establish its relation to the recent work of Callan and Symanzik.[10] We then review the program of Johnson, Baker, and Willey for the removal of infinities from electrodynamics. In Sec. III we show that the zero of the Gell-Mann–Low function must be an essential singularity and discuss the implications of this for the conventional

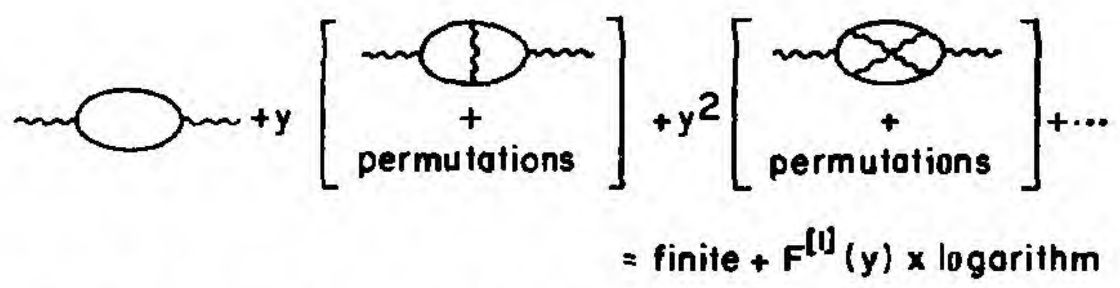


FIG. 1. Sum of single-fermion-loop vacuum-polarization diagrams which determines the function $F^{[1]}(y)$, with the dependence on the coupling constant y indicated explicitly. Throughout the paper we will adhere to the convention of using solid lines to denote fermions, wavy lines to denote photons.

eigenvalue condition on α_0 and for the asymptotic behavior of the renormalized electron propagator. In Sec. IV we introduce the idea of "loopwise" summation and show that, assuming the presence of the essential singularity, there is an asymptotically finite solution of electrodynamics in which α, rather than α_0, is fixed to be the infinite order zero. In Sec. V we motivate our conjecture that nature picks the solution in which α is fixed, and we suggest a possible connection of our work with a conjecture of Dyson[9] concerning singularities in electrodynamics at $\alpha = 0$. We also point out that our conjecture leads to a determination of α which is independent of the number of fundamental fermion species, and based on this fact, give a speculative argument justifying the neglect of the strong interactions in formulating our eigenvalue condition for α. In Appendix A we give a summary of notation, while in Appendix B we derive the Callan-Symanzik equations for massive photon (i.e., infrared-cutoff) spinor electrodynamics in an arbitrary covariant gauge, and briefly sketch the application of these equations to deriving the Johnson-Baker-Willey asymptotic form for the electron propagator.

II. REVIEW OF PREVIOUS WORK

We begin with a survey of the papers of Gell-Mann and Low, of Callan and Symanzik and of Johnson, Baker and Willey dealing with the short distance behavior of electrodynamics. Our particular aim will be to examine the underlying assumptions which these authors make and to discuss the connections between their approaches.

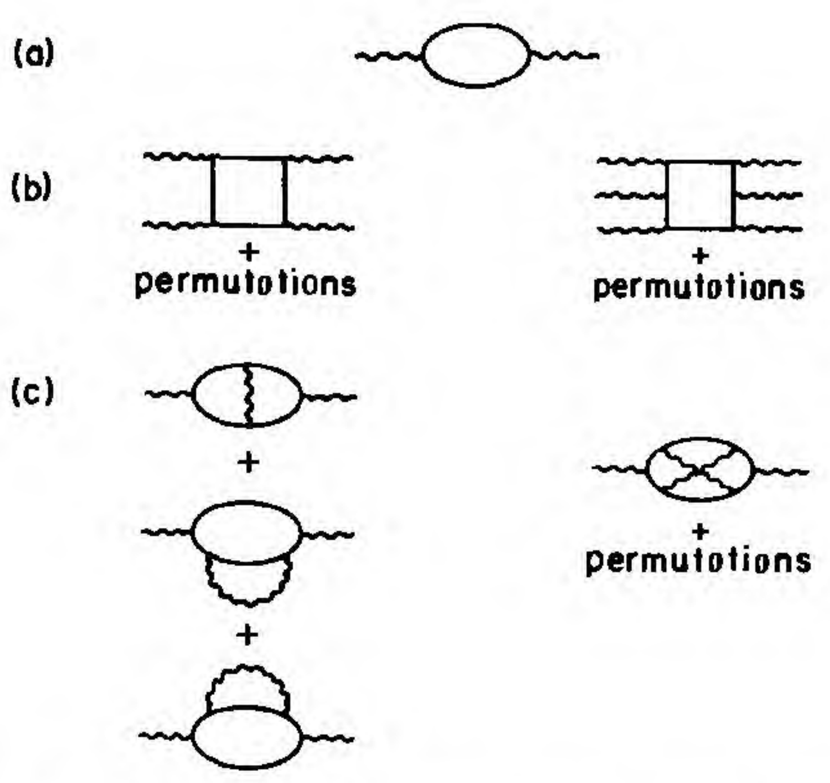


FIG. 2. (a) Lowest-order vacuum-polarization contribution to $\pi(q^2)_{\mu\nu}$. (b) Vacuum-polarization loops with four or more vertices. (c) Vacuum-polarization contributions to $\pi(q^2)$ which involve the loops with four or more vertices illustrated in (b).

A. Cutoff (Unrenormalized) and Renormalized Quantum Electrodynamics

In order for the renormalization constants and the unrenormalized propagators and vertex parts to be well-defined, it is necessary to introduce cutoffs. In addition to an ultraviolet cutoff Λ, we will eliminate infrared divergences by giving the photon a nonzero mass μ. The infrared cutoff will be needed for the derivation of the Callan-Symanzik equations for the electron propagator given in Appendix B. Where no infrared divergences are encountered, such as in the discussion of the asymptotic photon propagator which occupies the bulk of the paper, the photon mass μ will be set to zero. Specifically, we introduce the cutoffs as follows:

(i) The propagator for a bare photon of four-momentum q is given by

$$D_F^0(q)_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right)\frac{1}{q^2 - \mu_0{}^2}\,\frac{-\Lambda^2}{q^2 - \Lambda^2} + Z_3(\xi - 1)\frac{q_\mu q_\nu}{q^2}\,\frac{1}{q^2 - \mu^2}\,\frac{-\Lambda^2}{q^2 - \Lambda^2}, \quad (2)$$

with μ_0 the bare photon mass and ξ a guage parameter. The reason for the peculiar choice of the longitudinal term in Eq. (2) will become apparent very soon.

(ii) The lowest-order vacuum-polarization contribution to the photon proper self-energy $\pi(q^2)_{\mu\nu}$ comes from the fermion loop diagram illustrated in Fig. 2(a). We calculate this contribution in the following manner: First we impose gauge invariance to remove the quadratic divergence, and then we regulate the fermion loop, with fermion regulator mass Λ, to remove the logarithmic divergence.

(iii) All vacuum polarization loops with four or more vertices [see Fig. 2(b)] are calculated by imposing gauge invariance, which makes them finite. The requirement of gauge-invariant calculation of fermion loops, together with the photon propagator cutoff specified in (i), renders convergent the vacuum polarization contribution to $\pi(q^2)_{\mu\nu}$ of the type illustrated in Fig. 2(c). The photon propagator cutoff also makes finite all electron self-energy parts and vertex parts, so our cutoff procedure is sufficient to make the unrenormalized theory well defined.

We can now proceed to define renormalization constants and renormalized (i.e., Λ-independent in the limit $\Lambda\to\infty$) n-point functions. The renormalized electron propagator and electron-photon vertex are introduced in the standard[11] manner; we review in detail only the construction of the renormalized photon propagator. Since rules (ii)

and (iii) guarantee the gauge invariance of the photon proper self-energy part, we may write

$$\pi(q^2)_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right) q^2 \pi(q^2). \tag{3}$$

Letting α_b denote the canonical (bare) coupling and summing the series illustrated in Fig. 3 to get the full unrenormalized photon propagator $D_F'(q)_{\mu\nu}$, we get

$$D_F'(q)_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right) \times \frac{1}{q^2 - \mu_0^2 + \alpha_b q^2 \pi(q^2)[1 + O(q^2/\Lambda^2)]} \frac{-\Lambda^2}{q^2 - \Lambda^2} + Z_3(\xi - 1)\frac{q_\mu q_\nu}{q^2} \frac{1}{q^2 - \mu^2} \frac{-\Lambda^2}{q^2 - \Lambda^2}. \tag{4}$$

We fix the unrenormalized photon mass μ_0^2 by requiring Eq. (4) to have a pole at $q^2 = \mu^2$, i.e.,

$$\mu^2 - \mu_0^2 + \alpha_b \mu^2 \pi(\mu^2) = 0. \tag{5}$$

We then make the algebraic rearrangement

$$\begin{aligned} q^2 - \mu_0^2 + \alpha_b q^2 \pi(q^2) &= q^2 - \mu^2 + \alpha_b[q^2\pi(q^2) - \mu^2\pi(\mu^2)] \\ &= (q^2 - \mu^2)[1 + \alpha_b \pi(\mu^2)] + \alpha_b q^2[\pi(q^2) - \pi(\mu^2)] \\ &= Z_3^{-1}\{q^2 - \mu^2 + \alpha q^2[\pi(q^2) - \pi(\mu^2)]\}, \end{aligned} \tag{6}$$

which introduces the photon wave-function renormalization constant Z_3,

$$Z_3^{-1} = 1 + \alpha_b \pi(\mu^2), \tag{7a}$$

and the renormalized coupling constant α,

$$\alpha = \alpha_b Z_3 = \frac{\alpha_b}{1 + \alpha_b \pi(\mu^2)}. \tag{7b}$$

Comparing Eq. (7) with Eq. (5), we note that the photon bare mass can be reexpressed as

$$\mu_0^2 = Z_3^{-1}\mu^2, \tag{8}$$

indicating that it is not an independent renormalization constant. To get the full renormalized photon propagator, we multiply Eq. (4) by α_b and let the cutoff Λ become infinite, giving

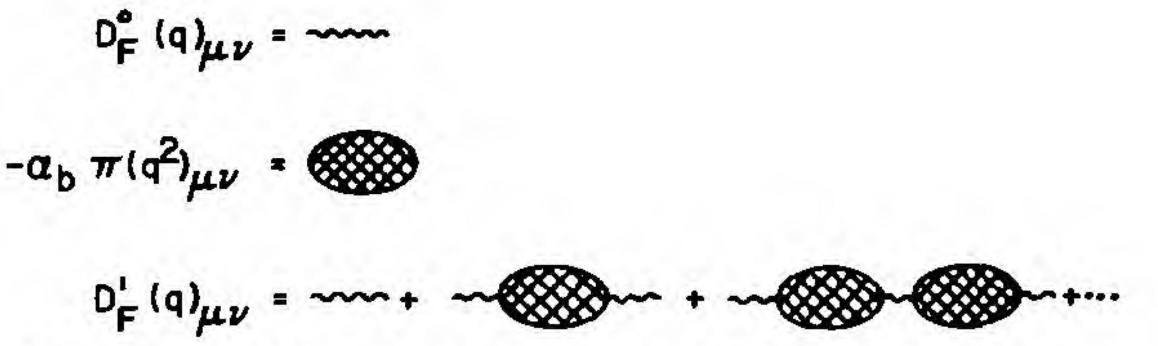


FIG. 3. Series which defines the full unrenormalized photon propagator $D_F'(q)_{\mu\nu}$.

$$\alpha \bar{D}_F'(q)_{\mu\nu} = \lim_{\Lambda\to\infty} \alpha_b D_F'(q)_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right)\frac{\alpha}{q^2 - \mu^2 + \alpha q^2 \pi_c(q^2)} + \alpha(\xi - 1)\frac{q_\mu q_\nu}{q^2}\frac{1}{q^2 - \mu^2}, \tag{9}$$

with

$$\pi_c(q^2) = \lim_{\Lambda\to\infty}[\pi(q^2) - \pi(\mu^2)]. \tag{10}$$

We can now see why the longitudinal part of the bare propagator had to be chosen as in Eq. (2): Because of the transversality of $\pi(q^2)_{\mu\nu}$, the longitudinal part of the full propagator [Eq. (4)] is the same as the longitudinal part of the bare propagator. Therefore, in order for the longitudinal part of the renormalized propagator to be finite, the longitudinal part of the bare propagator must become finite when multiplied by α_b. This dictates the overall factor of Z_3, and the use of μ^2 rather than μ_0^2 in the denominator. The fact that the gauge parameter ξ always occurs in the combination $(\xi - 1)Z_3$ will be of importance in the derivation of the Callan-Symanzik equations for the electron propagator given in Appendix B. On the other hand, the form of the longitudinal terms is irrelevant to the subsequent discussion of the Gell-Mann-Low and Callan-Symanzik equations for the photon propagator, because the photon proper self-energy is strictly gauge-invariant (rather than being merely gauge-covariant, as is the case for the electron propagator and the electron-photon vertex) and hence receives no contribution from the longitudinal terms.

To conclude this section, we state the specialization of Eqs. (7)-(10) to the case of massless-photon electrodynamics ($\mu^2 = \mu_0^2 = 0$). We have

$$\alpha \bar{D}_F'(q)_{\mu\nu} = \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right)\frac{\alpha d_c(-q^2/m^2, \alpha)}{q^2} + \alpha(\xi - 1)\frac{q_\mu q_\nu}{(q^2)^2}, \tag{11}$$

with

$$d_c(-q^2/m^2, \alpha) = [1 + \alpha\pi_c(q^2)]^{-1} \tag{12}$$

a dimensionless function which contains all the dynamical effects of vacuum polarization, and with $\pi_c(q^2)$ now given by

$$\pi_c(q^2) = \lim_{\Lambda\to\infty}[\pi(q^2) - \pi(0)]. \tag{13}$$

B. The Gell-Mann-Low Equation

We turn next to a review of the Gell-Mann-Low equation, which describes the asymptotic properties of the photon propagator. It will be useful to define an "asymptotic part" of the renormalized photon

propagator, which we denote by $\alpha d_c^\infty(-q^2/m^2, \alpha)$, in the following manner: We develop $\alpha d_c(-q^2/m^2, \alpha)$, in a perturbation expansion in powers of α and *in each order of perturbation theory* drop terms which vanish as $-q^2/m^2 \to \infty$, while retaining terms which are constant or which increase logarithmically.[12] The resulting sum of constant and logarithmic terms is the "asymptotic part" and clearly has the form

$$\alpha d_c^\infty(-q^2/m^2, \alpha) = q(\alpha) + p(\alpha)\ln(-q^2/m^2) + r(\alpha)\ln^2(-q^2/m^2) + \cdots . \tag{14}$$

Throughout the analysis which follows we will make the assumption that the nonasymptotic terms which we have neglected in each order of perturbation theory do not sum to give a result which dominates asymptotically over the logarithmic series in Eq. (14). That is, we assume that the asymptotic behavior of the "asymptotic part" αd_c^∞ correctly describes the asymptotic behavior of the exact propagator αd_c.[13]

To facilitate the derivation of the Gell-Mann–Low equation we introduce a notation which explicitly indicates the point where the subtraction in the photon proper self-energy is made. Thus, letting $x = -q^2/m^2$, we write

$$\begin{aligned} \alpha d_c[x, w, \alpha] &\equiv \alpha\{1 + \alpha(\pi[x] - \pi[w])\}^{-1}, \\ \pi[x] &\equiv \pi(-m^2 x) = \pi(q^2) . \end{aligned} \tag{15}$$

In terms of the new notation, the usual renormalized photon propagator is

$$d_c(x, \alpha) = d_c[x, 0, \alpha], \tag{16}$$

with the renormalized charge α given by

$$\alpha = \alpha d_c[0, 0, \alpha]. \tag{17}$$

Let us now imagine that, instead of making expansions in powers of the usual fine-structure constant α, we use as expansion parameter a new fine-structure constant α_w defined by

$$\alpha_w = \alpha d_c[w, 0, \alpha] = \alpha\{1 + \alpha(\pi[w] - \pi[0])\}^{-1}. \tag{18}$$

From the definition of Eq. (15), we may write

$$\begin{aligned} \alpha_w d_c[x, w, \alpha_w] &= \alpha_w\{1 + \alpha_w(\pi[x] - \pi[w])\}^{-1} \\ &= (\alpha_w^{-1} + \pi[x] - \pi[w])^{-1}, \end{aligned} \tag{19}$$

which on substitution of Eq. (18) becomes

$$\begin{aligned} \alpha_w d_c[x, w, \alpha_w] &= \{\alpha^{-1} + \pi[y] - \pi[0] + \pi[x] - \pi[w]\}^{-1} \\ &= \alpha d_c[x, 0, \alpha] = \alpha d_c(x, \alpha) . \end{aligned} \tag{20}$$

On the right-hand side we have the usual photon propagator, which involves a subtraction at the nonasymptotic point zero; Eq. (20) states that this can be reexpressed in terms of the new charge α_w and the photon propagator subtracted at w, with no further reference to the point zero.

Let us now let x and w both become large. According to our earlier discussion, the right-hand side of Eq. (20) becomes the logarithmic series $\alpha d_c^\infty(x, \alpha)$. For the left-hand side, we introduce the asymptotic assumption that the only dependence on x and w, when both are large, is through the ratio x/w. An equivalent statement of the asymptotic assumption is that when $x = -q^2/m^2$ and $w = -q'^2/m^2$ are both large, the quantity $\alpha_w d_c[x, y, \alpha_w]$, regarded as a power series in α_w, becomes independent of the electron mass m.[14] This assumption can actually be justified order-by-order in perturbation theory, either by using the analysis of Callan and Symanzik (see below) or by invoking the theorem on cancellation of infrared singularities of Kinoshita[15] and Lee and Nauenberg.[15] Equation (20) now becomes

$$\alpha_w D[x/w, \alpha_w] = \alpha d_c^\infty(x, \alpha), \tag{21}$$

where, since w is large, we may rewrite Eq. (18) for α_w as

$$\begin{aligned} \alpha_w &= \alpha d_c^\infty(w, \alpha) \\ &= q(\alpha) + p(\alpha)\ln w + r(\alpha)\ln^2 w + \cdots . \end{aligned} \tag{22}$$

Equation (21) gives a functional relation for d_c^∞, which may be rewritten in a more useful form as follows: We introduce the function $\psi(z)$ by the definition

$$\psi(z) = \frac{\partial}{\partial v} zD[v, z]\bigg|_{v=1} . \tag{23}$$

Differentiating Eq. (21) with respect to x, and then setting $x = w$, we get the differential equation

$$\frac{1}{w}\psi(\alpha_w) = \frac{d\alpha_w}{dw} . \tag{24a}$$

Rewriting this as

$$\frac{dw}{w} = \frac{d\alpha_w}{\psi(\alpha_w)} , \tag{24b}$$

integrating with respect to w from 1 to x and using the boundary condition

$$\alpha_w|_{w=1} = q(\alpha) = \alpha d_c^\infty(1, \alpha) \tag{25}$$

we get finally the Gell-Mann–Low equation,

$$\ln x = \int_{q(\alpha)}^{\alpha d_c^\infty(x,\alpha)} \frac{dz}{\psi(z)} . \tag{26}$$

It is also useful to have the inversion formula relating the coefficients $q(\alpha), p(\alpha), \ldots,$ in the logarithmic expansion of Eq. (22) to the Gell-Mann–Low function $\psi(z)$. To get this, we write $z = \alpha_x = \alpha d_c^\infty(x, \alpha)$ and make a Taylor expansion of z with respect to $\ln x$,

$$z = \sum_{n=0}^{\infty} \frac{(\ln x)^n}{n!} \frac{d^n}{d(\ln x)^n} z \bigg|_{\ln x = 0} \tag{27a}$$

According to Eq. (24), the derivative $d/d(\ln x)$ may be rewritten as

$$\frac{d}{d(\ln x)} = \frac{dz}{d(\ln x)} \frac{d}{dz} = \psi(z) \frac{d}{dz}, \tag{27b}$$

giving the desired formula[6]

$$\alpha d_c^\infty(x, \alpha) = \sum_{n=0}^{\infty} \frac{(\ln x)^n}{n!} \left\{ \left[\psi(z) \frac{d}{dz} \right]^n z \right\} \bigg|_{z = q(\alpha)}. \tag{28}$$

The function $\psi(z)$ appearing in these formulas is not explicitly known beyond its expansion to sixth order of perturbation theory, which is[18]

$$\psi(z) = z \left(\frac{z}{3\pi} + \frac{z^2}{4\pi^2} + \frac{z^3}{8\pi^3} \left[\tfrac{8}{3} \zeta(3) - \tfrac{101}{36} \right] + \cdots \right), \tag{29}$$

with $\zeta(3)$ the Riemann ζ function.

As Gell-Mann and Low have shown, Eq. (26) provides a powerful tool for analyzing the asymptotic behavior of the photon propagator, and leads one naturally to distinguish the following two possibilities: (a) The integral $\int dz/\psi(z)$ in Eq. (26) does not diverge until the upper limit becomes infinite. In this case $\alpha d_c^\infty(x, \alpha) \to \infty$ as $x \to \infty$, and so the photon propagator is asymptotically divergent. (b) For some finite value $z = \alpha_0$, the function $\psi(z)$ develops a sufficiently strong zero for $\int^{\alpha_0} dz/\psi(z)$ to diverge. In this case $\alpha d_c^\infty(x, \alpha) \to \alpha_0$ as $x \to \infty$ and the photon propagator is asymptotically finite. We will restrict our attention from here on exclusively to case (b), for which, as noted in the Introduction, we may write

$$\alpha d_c^\infty(x, \alpha) = \alpha_0 + h(x, \alpha), \qquad \lim_{x \to \infty} h(x, \alpha) = 0. \tag{30}$$

Within the category of case (b), we wish to further distinguish between two different types of possible asymptotic behavior of the theory:

Type 1. The physical fine-structure constant α is equal to the particular value α_1 which satisfies

$$q(\alpha_1) = \alpha_0. \tag{31a}$$

According to Eq. (21), the coefficient of $(\ln x)^n$ with $n \geq 1$ is then

$$\left\{ \left[\psi(z) \frac{d}{dz} \right]^n z \right\} \bigg|_{z = q(\alpha)} = \psi(q(\alpha)) \left\{ \frac{d}{dz} \left[\psi(z) \frac{d}{dz} \right]^{n-1} z \right\} \bigg|_{z = q(\alpha)} = \psi(\alpha_0)\{\cdots\} = 0; \tag{31b}$$

and the logarithmic series reduces to its constant term alone,

$$\alpha d_c^\infty(x, \alpha) = \alpha_0. \tag{31c}$$

In this case the Gell-Mann-Low equation degenerates to an integral over an interval of vanishing size located at the point where ψ^{-1} is infinite.

Type 2. The physical fine-structure constant α differs from α_1. The coefficients of the logarithmic terms in Eq. (28) then do not vanish and $\alpha d_c^\infty(x, \alpha)$ is a nontrivial function of x which approaches α_0 in the limit as $x \to \infty$. In this case the Gell-Mann-Low equation is nondegenerate, with the integral extending over an interval of finite size, and α is an undetermined parameter.

Clearly, as far as behavior of the photon propagator is concerned, the more general class of asymptotically finite solutions with type-2 behavior is just as satisfactory physically as the solution with type-1 behavior. (We will find additional evidence for this statement when we study the asymptotic electron propagator below.) Hence following Gell-Mann and Low, we conclude that requiring asymptotic finiteness of the renormalized photon propagator fixes α_0, but does not determine the fine-structure constant α.

To conclude this discussion of the Gell-Mann-Low equation we give a simple, concrete illustration of type-2 asymptotic behavior. Let us make the customary assumption that $\psi(z)$ is regular and vanishes with a simple zero and negative slope at $z = \alpha_0$. We ignore the fact that $\psi(z)$ also vanishes for small z and replace ψ by a linear approximation over the integration interval in the Gell-Mann-Low equation,

$$\psi(z) \approx \psi'(\alpha_0)(\alpha_0 - z), \quad \psi'(\alpha_0) < 0. \tag{32}$$

Then Eq. (26) can be immediately integrated to give

$$\ln x = \frac{1}{\psi'(\alpha_0)} \ln \left[\frac{\alpha d_c^\infty(x, \alpha) - \alpha_0}{q(\alpha) - \alpha_0} \right], \tag{33}$$

which can be rewritten as

$$\alpha d_c^\infty(x, \alpha) = \alpha_0 + [q(\alpha) - \alpha_0] x^{\psi'(\alpha_0)} = q(\alpha) + [q(\alpha) - \alpha_0] \sum_{n=1}^{\infty} \frac{\psi'(\alpha_0)^n (\ln x)^n}{n!}. \tag{34}$$

We see that the logarithmic series for $\alpha d_c^\infty(x, \alpha)$ is nontrivial, with all powers of $\ln x$ present, but that it sums to a function which approaches α_0 asymptotically. The nonasymptotic piece h, which is given in our example by

$$h(x, \alpha) = [q(\alpha) - \alpha_0] x^{\psi'(\alpha_0)}, \tag{35}$$

vanishes asymptotically as a power of x independently of the value of α.

C. The Callan-Symanzik Equation

A very powerful and elegant method for studying the asymptotic behavior of renormalized perturbation theory has recently been developed by Callan and Symanzik.[10] We briefly review here the derivation of the Callan-Symanzik equation for the renormalized photon propagator,[17] and indicate its connection with the Gell-Mann-Low equation discussed above. Our starting point is the formula relating the renormalized and unrenormalized photon propagators,

$$d_c(-q^2/m^2,\alpha)^{-1}=Z_3(\Lambda^2/m^2,\alpha)[1+\alpha_b\pi(q^2)], \tag{36}$$

where we have explicitly indicated the cutoff dependence of Z_3. Since the photon propagator is gauge invariant, the quantities appearing in Eq. (36) have no dependence on the gauge parameter ξ. Let us now vary the physical electron mass m, with the canonical (bare) coupling α_b and the cutoff Λ held fixed. Under this variation the bare electron mass m_0 and the physical coupling α both change, since the renormalization conditions give both of them an implicit dependence on m. Thus, insofar as d_c and Z_3 are concerned, variation of m is described by

$$m\frac{d}{dm}=m\frac{\partial}{\partial m}+m\frac{d\alpha}{\partial m}\frac{\partial}{\partial\alpha}, \tag{37}$$

while for the bare propagator $1+\alpha_b\pi(q^2)$, which depends on m only through m_0, the mass variation is described by

$$m\frac{d}{dm}=m\frac{dm_0}{dm}\frac{\partial}{\partial m_0}. \tag{38}$$

Equating the mass variations of the left- and right-hand sides of Eq. (36) gives

$$\begin{aligned}\left(m\frac{\partial}{\partial m}+m\frac{d\alpha}{dm}\frac{\partial}{\partial\alpha}\right)d_c^{-1}&=m\frac{d}{dm}d_c^{-1}\\&=m\frac{d}{dm}Z_3[1+\alpha_b\pi]+Z_3m\frac{d}{dm}[1+\alpha_b\pi]\\&=Z_3^{-1}m\frac{d}{dm}Z_3d_c^{-1}+\alpha m\frac{dm_0}{dm}\frac{\partial}{\partial m_0}\pi.\end{aligned} \tag{39}$$

The term $\partial\pi/\partial m_0$ on the right-hand side of Eq. (39) is simply interpreted as a photon-photon-scalar vertex part, with the scalar current carrying zero four-momentum. It can be shown[18] that this vertex part is made finite by multiplication by the renormalization constant m_0, and so we can write

$$m_0\frac{\partial}{\partial m_0}\pi=\bar{\Gamma}_{\gamma\gamma S}(q^2/m^2,\alpha). \tag{40}$$

It can be further shown[10,18] that the quantities $\beta(\alpha)$ and $\delta(\alpha)$ defined by

$$\begin{aligned}\beta(\alpha)&=Z_3^{-1}m\frac{d}{dm}Z_3,\\1+\delta(\alpha)&=m_0^{-1}m\frac{d}{dm}m_0\end{aligned} \tag{41}$$

are cutoff-independent and therefore, as indicated, are functions of α alone. Finally, we can relate the quantity $m\,d\alpha/dm$ appearing on the left-hand side of Eq. (39) to $\beta(\alpha)$, as follows:

$$\begin{aligned}m\frac{d\alpha}{dm}&=m\frac{d}{dm}(\alpha_bZ_3)=\alpha_bm\frac{d}{dm}Z_3\\&=\alpha Z_3^{-1}m\frac{d}{dm}Z_3=\alpha\beta(\alpha).\end{aligned} \tag{42}$$

Putting everything together we get the Callan-Symanzik equation for the photon propagator,

$$\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]d_c(-q^2/m^2,\alpha)^{-1}=\alpha[1+\delta(\alpha)]\bar{\Gamma}_{\gamma\gamma S}(q^2/m^2,\alpha). \tag{43}$$

Let us now let $-q^2/m^2$ become infinite. Order by order in perturbation theory, the left-hand side of Eq. (43) becomes

$$\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]d_c^{\infty}(-q^2/m^2,\alpha)^{-1}, \tag{44}$$

while a simple application of Weinberg's theorem[19] shows that, again order by order in perturbation theory, the right-hand side of Eq. (43) vanishes. So we learn that d_c^{∞} satisfies the differential equation

$$\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]d_c^{\infty}(-q^2/m^2,\alpha)^{-1}=0. \tag{45a}$$

Interestingly, when we substitute Eq. (37) for md/dm into Eq. (42), we learn[17] that $Z_3(\Lambda^2/m^2,\alpha)$ satisfies a differential equation identical in form to Eq. (45a),

$$\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]Z_3(\Lambda^2/m^2,\alpha)=0. \tag{45b}$$

For the subsequent discussion, it will be useful to reexpress Eq. (45a) as a differential equation for d_c^∞,

$$\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}+1\right)\right]d_c^\infty(-q^2/m^2,\alpha)=0. \qquad (46)$$

We will now show that Eq. (46) is completely equivalent to Eq. (26), the Gell-Mann–Low equation. Letting x, as before denote $-q^2/m^2$, we rewrite Eq. (46) in the form

$$\left[-2x\frac{\partial}{\partial x}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}\right]\alpha d_c^\infty(x,\alpha)=0. \qquad (47)$$

This equation has the integral

$$\alpha d_c^\infty(x,\alpha)=\Phi^{-1}\left[\ln x+\int_c^\alpha\frac{2dz}{z\beta(z)}\right], \qquad (48)$$

with the function Φ determined by the $x=1$ boundary condition

$$\Phi[\alpha d_c^\infty(1,\alpha)]=\Phi[q(\alpha)]=\int_c^\alpha\frac{2dz}{z\beta(z)} \qquad (49)$$

and with c an arbitrary constant of integration. (The presence of c merely reflects the freedom of changing Φ by an arbitrary additive constant.) Inverting Eq. (48), we thus can write

$$\ln x=\Phi[\alpha d_c^\infty(x,\alpha)]-\Phi[q(\alpha)] \qquad (50)$$

which, if we write $\Phi[u]$ in the integral form

$$\Phi[u]=\int_{c'}^u\frac{dz}{\psi(z)},\quad \psi(z)=[\Phi'(z)]^{-1}, \qquad (51)$$

can be further recast as

$$\ln x=\int_{q(\alpha)}^{\alpha d_c^\infty(x,\alpha)}\frac{dz}{\psi(z)}, \qquad (52)$$

which is just the Gell-Mann–Low equation. Clearly, the derivation which we have just given does not involve the asymptotic assumption made in the discussion immediately following Eq. (20); in effect, the Callan-Symanzik route to the Gell-Mann–Low equation replaces a statement about infrared behavior (m independence of $\alpha_w d_c[x,w,\alpha_w]$ as $m\to 0$) with a statement about ultraviolet behavior (asymptotic vanishing of $\bar{\Gamma}_{\gamma\gamma S}$) which can be justified in perturbation theory by the use of Weinberg's theorem.

Comparing Eq. (51) with Eq. (49), we can read off the following functional relationship between the Callan-Symanzik function $\beta(\alpha)$ and the Gell-Mann–Low function $\psi(z)$,

$$\beta(\alpha)=\frac{2\psi(q(\alpha))}{\alpha q'(\alpha)}. \qquad (53)$$

Thus, in the case of type-1 asymptotic behavior, where $\alpha=\alpha_1$, we have $\beta(\alpha)=0$. Equation (47) then reduces to

$$x\frac{\partial}{\partial x}\alpha d_c^\infty(x,\alpha)=0, \qquad (54)$$

which has, as expected, the integral

$$\alpha d_c^\infty(x,\alpha)=\alpha_0. \qquad (55)$$

Similarly, Eq. (45b) tells us that when $\beta(\alpha)=0$, the photon wave-function renormalization Z_3 is cutoff-independent. As shown in Appendix B, the Callan-Symanzik equation for the renormalized electron propagator also has the function $\beta(\alpha)$ as coefficient of the $\partial/\partial\alpha$ term. Consequently, in the case of type-1 asymptotic behavior this equation also simplifies, and leads, by a simple argument,[18] to an elementary scaling form for the asymptotic electron propagator. Clearly, in the case of type-2 asymptotic behavior we have $\beta(\alpha)\neq 0$ and must deal with the Callan-Symanzik equations in their full complexity. Even so, we find in Appendix B that the scaling form for the asymptotic electron propagator still holds, again indicating, as asserted above, that there is no reason for favoring the type-1 solution over the more general class of asymptotically finite solutions with type-2 asymptotic behavior.

D. The Johnson-Baker-Willey Program

We conclude our review by surveying the recent work of Johnson, Baker, and Willey (JBW) dealing with the asymptotic properties of electrodynamics. As noted in Sec. I, this work has led to two principal results: a simplified form of the Gell-Mann–Low eigenvalue condition for α_0, and a demonstration that if the eigenvalue condition is satisfied, then the electron bare mass m_0 and wave-function renormalization constant Z_2 can be finite. We discuss these two aspects in turn.

1. *Simplified Eigenvalue Condition*

A key ingredient in the JBW formulation of the eigenvalue condition is the use of "vacuum-polarization-insertion-wise" summation of the photon proper self-energy part π. The basic idea is to first write down a modified skeleton expansion for the photon proper self-energy in which all diagrams with internal vacuum polarization insertions are omitted. Some typical diagrams which appear in this expansion are illustrated in Fig. 4; note that they still contain internal electron self-energy and electron-photon vertex parts. The next step is to replace all internal photon lines appearing in the expansion by full renormalized photon propagators $\alpha\bar{D}_F'(q)_{\mu\nu}$ (we indicate explicitly the coupling constant α associated with the ends of the photon line).

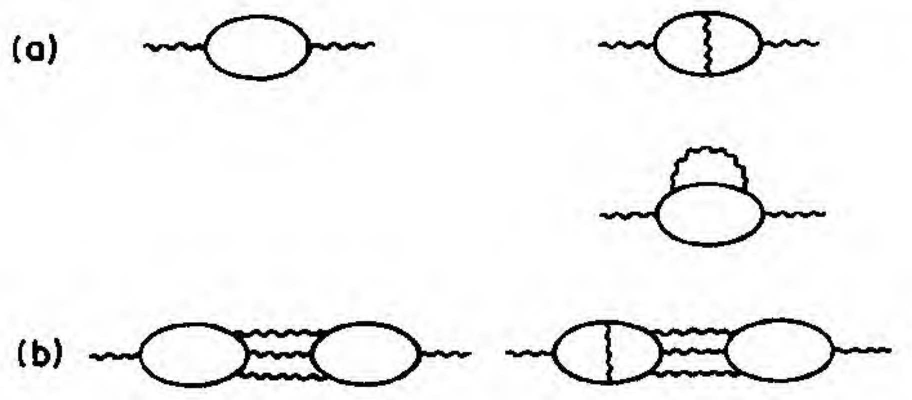


FIG. 4. Typical diagrams which appear in the modified skeleton expansion for the photon proper self-energy part π. All proper diagrams are included which do not have internal photon self-energy insertions. Diagrams (a) have a single fermion loop, while those labeled (b) contain two or more fermion loops.

This recipe leads to a "vacuum-polarization-insertion-wise" summed expression for the photon proper self-energy π, which, it is easy to see, correctly includes all of the relevant Feynman diagrams.

We next introduce the assumption that the renormalized photon propagator is asymptotically finite, which allows us to write it in the form

$$\alpha\tilde{D}_F'(q)_{\mu\nu}=\left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right)\frac{1}{q^2}[\alpha_0+h(-q^2/m^2,\alpha)] + \alpha(\xi-1)\frac{q_\mu q_\nu}{(q^2)^2}, \tag{56}$$

with h vanishing asymptotically. In order for this assumption to be self-consistent, we require that the renormalized photon proper self-energy part π_c, obtained by inserting Eq. (56) in the skeleton expansion as outlined above, must itself be asymptotically finite. That is, no powers of $\ln(-q^2/m^2)$ may be present in the asymptotic behavior of π_c. To determine the asymptotic properties of π_c, we must consider each contributing graph (or more exactly, each gauge-invariant set of graphs related by permutation of photon vertices) and examine the convergence properties of both the over-all integration, involving all lines in the graph, and the subintegrations involving subsets of these lines. In doing this we maintain our "vacuum-polarization-insertion-wise" summation by treating internal photon propagators as complete entities, described by Eq. (56), rather than also breaking these up into contributing graphs. By our assumption of Eq. (56), the internal photon propagators cannot give rise to any logarithmic terms. Subintegrations associated with electron-self-energy and electron-photon vertex parts also lead to no logarithms, because as shown in Sec. II D 2, there is a gauge (the Landau gauge) which makes these asymptotically finite. It can be shown[7,18] that there are no other troublesome subintegrations; hence logarithmic behavior of a graph contributing to π_c can only be associated with the over-all integration involving all lines in the graph. Referring to Eq. (56), we see that each internal photon line in the over-all integration contributes two parts, a part proportional to α_0 and a part proportional to h. As we have seen in Eqs. (32)–(35) above, the conventional assumptions about the nature of the zero of $\psi(z)$ imply that h decreases as a power of $-q^2/m^2$ as $-q^2/m^2\to\infty$. As a result, any contribution to the over-all integration involving one or more factors h converges, and leads to an asymptotically finite contribution to π_c. Hence the asymptotically logarithmic part of π_c is correctly obtained by neglecting h in each internal photon propagator, so that Eq. (56) becomes

$$\alpha\tilde{D}_F'(q)_{\mu\nu}=-g_{\mu\nu}\frac{\alpha_0}{q^2}+\text{gauge term}. \tag{57}$$

Thus, we are led to a simplified model for π_c (the so-called JBW model) in which no internal photon self-energy insertions appear; all internal photons are described by free propagators coupling with the asymptotic coupling strength α_0. An analysis[7,18] of the asymptotic behavior of π_c in this model shows that a single logarithm is present (corresponding to the fact that a single subtraction suffices to make the over-all integration converge), so we get finally

$$\pi_c \underset{-q^2/m^2\to\infty}{\sim} g(\alpha_0)+f(\alpha_0)\ln(-q^2/m^2). \tag{58}$$

Self-consistency of the assumption of asymptotic finiteness of π_c now requires

$$f(\alpha_0)=0, \tag{59}$$

which is the JBW form of the eigenvalue condition. Let us reiterate that Eq. (59) does not involve all vacuum polarization graphs [as does the Gell-Mann–Low eigenvalue condition $\psi(\alpha_0)=0$] but rather only the restricted class, illustrated in Fig. 4, which have no internal photon self-energy insertions.

Implicit in the derivation of Eq. (59) are rather stringent convergence assumptions. These arise because the argument leading to Eq. (59) involves replacing the limit of an infinite sum [the exact $\pi_c(q^2)$ is an infinite sum of skeleton graphs with photon self-energy insertions] by the sum of the limits of the individual terms. [Eq. (58) is the sum of skeletons with the photon self-energy insertions replaced by their asymptotic limits.] A necessary, but by no means sufficient, condition for the interchange of limit with sum to be valid is that the resulting series $f(\alpha_0)$ be convergent. This fact will be of importance in the discussion of "loopwise" summation given in Sec. IV below.

In their recent papers,[7] Baker and Johnson have extended in two respects the treatment of the eigen-

value equation sketched above: First, they have shown that $f(\alpha_0)=0$ implies that α_0 is also a zero of the Gell-Mann–Low function $\psi(y)$, and secondly, they have shown (again assuming "vacuum-polarization-insertion-wise" summation) that Eq. (59) can be replaced by the much simpler eigenvalue condition

$$F^{[1]}(\alpha_0)=0\,, \tag{60}$$

where $F^{[1]}(y)$ is the single fermion loop part of $f(y)$ introduced in Sec. I. The first assertion is proved by an argument (which we omit) based on properties of the modified skeleton expansion, showing that $\psi(y)$ and $f(y)$ are functionally related,

$$\begin{aligned}\psi(y)&=\sum_{n=1}^{\infty}[f(y)]^n c_n(y)\\ &=f(y)c_1(y)+f^2(y)c_2(y)+\cdots.\end{aligned} \tag{61}$$

Hence a zero of f is necessarily a zero of ψ. The second assertion follows from a simple argument based on the Federbush-Johnson theorem; we give details in the case, since the results are central to the discussion of Sec. III below. We assume that the Gell-Mann-Low eigenvalue equation $\psi(y)=0$ has a solution $y=\alpha_0$, so that the renormalized photon propagator takes the form of Eq. (1). If we now let the electron mass m approach zero, we learn from Eq. (1) that the renormalized photon propagator d_c approaches its asymptotic value α_0 for any $q^2\neq 0$. This means that *in a theory of massless spin-$\frac{1}{2}$ electrodynamics satisfying the eigenvalue condition, the full renormalized photon propagator is exactly equal to the free photon propagator*, with coupling constant α_0. Consequently, the absorptive part of the photon proper self-energy vanishes; i.e., we have

$$\langle 0|j_\mu(x)j_\nu(y)|0\rangle=0\,, \tag{62}$$

where j_μ is the electromagnetic current operator. By exploiting positivity of the absorptive part of the full photon propagator, Federbush and Johnson[8,20] have shown that the vanishing of the two-point function in Eq. (62) implies that $j_\mu(x)$ annihilates the vacuum, and hence the general $2n$-point current correlation function vanishes as well,

$$\langle 0|T[j_{\mu_1}(x_1)j_{\mu_2}(x_2)\cdots j_{\mu_{2n}}(x_{2n})]|0\rangle=0,\quad n\geq 2\,. \tag{63}$$

Equation (63) is the essential tool which allows us to simplify the eigenvalue condition. Let us take the difference between the photon self-energy part π_c evaluated at four-momenta q^2 and q'^2. Since the full photon propagator is equal to the free photon propagator in the massless theory, this difference may be calculated from the skeleton diagrams of Fig. 4, and according to Eq. (58) is given by

$$\pi_c(q^2)-\pi_c(q'^2)=f(\alpha_0)\ln(q^2/q'^2)\,. \tag{64}$$

The contributions to Eq. (64) may be divided into two basic types: those containing a single closed fermion loop [Fig. 4(a)] and those containing two or more closed fermion loops [Fig. 4(b)]. The sum of contributions of the second type can be recast as a sum involving current correlation functions which have been linked together by photon lines, and therefore vanishes by Eq. (63). Thus, the vanishing of the logarithmic term in Eq. (64) implies that the sum of contributions of the first type must vanish by itself, which gives the simplified eigenvalue condition

$$F^{[1]}(\alpha_0)=0\,. \tag{65}$$

Clearly, the same argument applied to Eq. (63) shows that *the sum of single closed fermion loop contributions to the general $2n$-point current correlation function ($n\geq 2$) must vanish by itself when the coupling is α_0 and the fermion is massless*, a result which will be of great utility in the next section. We stress in closing that the powerful results which we have just described are consequences of positivity of the spectral function of the photon propagator. In particular, since the single closed fermion loop contributions to π_c do not by themselves have a positive spectral function, the methods which we have used cannot be used to prove the converse result that a zero of $F^{[1]}[y]$ is necessarily a zero of $f(y)$ and $\psi(y)$.[21]

2. *Asymptotic Electron Propagator and Finiteness of Z_2 and m_0*

To analyze the asymptotic electron propagator, JBW employ the simplified model described above, in which the asymptotically vanishing part h of the photon propagator is neglected. Each internal photon is thus represented by a free propagator, coupling with the asymptotic coupling strength α_0. In this model it is straightforward to determine the asymptotic behavior of the renormalized electron propagator and of the renormalization constants Z_2 and m_0, either by using renormalization group methods[7] or by use of the Callan-Symanzik equation,[18] with the results

$$\begin{aligned}\tilde S_F'(p)^{-1}\underset{p\to\infty}{\sim}\;&F_1(\alpha_1)C_1(\mu^2/m^2,\alpha_1)\left(-\frac{p^2}{m^2}\right)^{\gamma(\alpha_1)/2}\\ &\times\left[\not p-mF_2(\alpha_1)C_2(\mu^2/m^2,\alpha_1)\left(-\frac{p^2}{m^2}\right)^{-\delta(\alpha_1)/2}\right],\end{aligned} \tag{66}$$

$$Z_2=C_1(\mu^2/m^2,\alpha_1)\left(\frac{\Lambda^2}{m^2}\right)^{\gamma(\alpha_1)/2},$$

$$m_0=C_2(\mu^2/m^2,\alpha_1)m\left(\frac{\Lambda^2}{m^2}\right)^{-\delta(\alpha_1)/2}.$$

In writing Eq. (66) we have used the fact that in the

JBW model the mapping $q(\alpha)$ is effectively the unit mapping $q(\alpha)=\alpha$, and so Eq. (30) tells us that

$$\alpha_0 = q(\alpha_1) = \alpha_1 . \tag{67a}$$

The function $\delta(\alpha_1)$ is defined in Eq. (41), while the definition of $\gamma(\alpha_1)$ is given in Appendix B. The transformation properties of Eq. (17) under changes in the gauge parameter ξ can be explicitly worked out,[18] and for the exponents γ and δ we find (primed quantities refer to gauge parameter ξ', unprimed to gauge parameter ξ)

$$\begin{aligned}\gamma' - \gamma &= \frac{\alpha_1}{2\pi}(\xi' - \xi),\\ \delta' - \delta &= 0 .\end{aligned} \tag{67b}$$

Thus, if we choose ξ' to satisfy

$$(\alpha_1/2\pi)(\xi' - \xi) + \gamma(\alpha_1, \xi) = 0,$$

then we have $\gamma' \equiv \gamma(\alpha_1, \xi') = 0$ and the electron wave function renormalization Z_2' remains finite as $\Lambda \to \infty$. Furthermore, if $\delta(\alpha_1) > 0$, the electron bare mass m_0 vanishes in the limit of infinite cutoff, indicating that the physical mass of the electron is entirely electromagnetic in origin. The apparent logarithmic divergence of m_0 in perturbation theory results only when

$$m_0 = C_2(\mu^2/m^2, \alpha_1) m \exp[-\tfrac{1}{2}\delta(\alpha_1)\ln(\Lambda^2/m^2)] \tag{68}$$

is expanded in a power series in $\alpha_1 = \alpha_0$ and illegally truncated at a finite order. Thus, in the model with the photon propagator replaced by its finite asymptotic part, all perturbation theory infinities can be eliminated, provided only that $\delta(\alpha_1) > 0$.

A little caution is required, however, in applying the results of the JBW model to the full theory, where the photon propagator contains the nonasymptotic piece h in addition to the asymptotic part α_0. Because the renormalization counterterms which are subtracted in going from the unrenormalized to the renormalized electron propagator are evaluated at the nonasymptotic four-momentum $\not p = m$, it is easy to see that h makes a nonvanishing contribution to the asymptotic renormalized electron propagator. Thus Eq. (66) does not necessarily apply to the full theory. In Appendix B we analyze the effect of h on the asymptotic behavior of $\tilde{S}_F'(p)^{-1}$. Assuming that h vanishes asymptotically as a power of $-q^2/m^2$, we find that the *form* of Eq. (66) and of the exponents γ and δ are unaltered, all of the effects of h being confined to changes in the constants C_1 and C_2.[22] Hence, when h vanishes as a power, the conclusions obtained from the JBW model regarding the finiteness of Z_2 and m_0 apply to the full theory as well.

III. THE ESSENTIAL SINGULARITY AND ITS CONSEQUENCES

We continue in the present section to work with the "vacuum-polarization-insertion-wise" summation scheme described above in Sec. II D 1. We show that the argument leading to the simplified eigenvalue condition of Eq. (65) has the further implication that $F^{[1]}(y)$ vanishes at $y=\alpha_0$ with a zero of infinite order, i.e., an essential singularity. We find that as a result, the nonasymptotic piece h of the photon propagator vanishes asymptotically much more slowly than any power of $-q^2/m^2$, and we discuss consequences of this both for the eigenvalue condition and for the asymptotic behavior of the electron propagator.

A. Existence of an Essential Singularity

Since our argument makes extensive use of the properties of the single-fermion-loop $2n$-point functions, we begin by introducing a compact notation for these. Let us denote the sum of single-fermion-loop contributions to the photon proper self-energy by

$$\pi_c^{[1]}(q^2; m, y), \tag{69}$$

where we have explicitly indicated the dependence on the fermion mass m and on the coupling constant y. The series of diagrams defining $\pi_c^{[1]}$ has, of course, already been exhibited in Fig. 1. According to the results of Sec. II D, when $-q^2/m^2$ approaches infinity $\pi_c^{[1]}$ has the asymptotic behavior

$$\pi_c^{[1]}(q^2; m, y) = G^{[1]}(y) + F^{[1]}(y)\ln(-q^2/m^2) + \text{vanishing terms}, \tag{70}$$

and the assumption that the Gell-Mann-Low function ψ vanishes at $y=\alpha_0$ implies that the coefficient of the logarithm in Eq. (70) also vanishes for this value of the coupling,

$$F^{[1]}(\alpha_0) = 0 . \tag{71}$$

Let us next denote the sum of single-fermion-loop contributions to the general $2n$-point current correlation function ($n \geq 2$) by

$$\begin{aligned}T_{2n}^{[1]} &= T^{[1]}_{\mu_1 \cdots \mu_{2n}}(q_1, \ldots, q_{2n}; m, y),\\ q_1 + \cdots + q_{2n} &= 0;\end{aligned} \tag{72}$$

the series of diagrams defining $T_{2n}^{[1]}$ is shown in Fig. 5. In each order in the power series expansion in y, all distinct permutations of external and internal photon vertices are included in $T_{2n}^{[1]}$. As a result, $T_{2n}^{[1]}$ is independent of the gauge parameter ξ appearing in the internal photon propagators and satisfies current conservation with respect to the external photon indices,

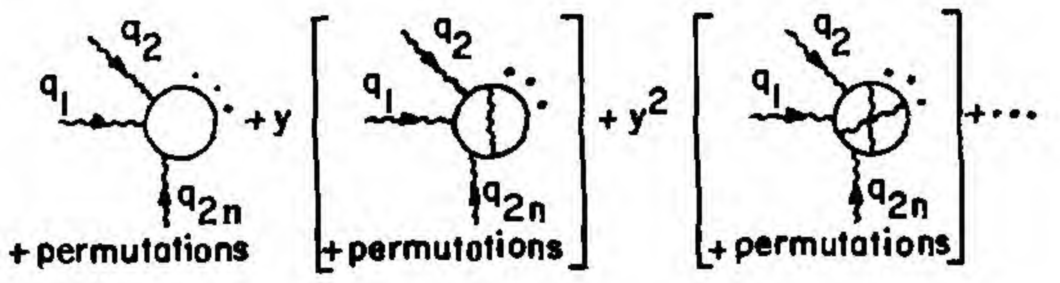


FIG. 5. Sum of single-fermion-loop diagrams which defines the $2n$-point function $T_{2n}^{[1]}$ appearing in Eq. (72), with the dependence on the coupling constant y indicated explicitly.

$$q_1^{\mu_1} T^{[1]}_{\mu_1 \cdots \mu_{2n}} = q_2^{\mu_2} T^{[1]}_{\mu_1 \cdots \mu_{2n}} = \cdots = q_{2n}^{\mu_{2n}} T^{[1]}_{\mu_1 \cdots \mu_{2n}} = 0 . \tag{73}$$

As was shown in Sec. II D, when the fermion mass m is zero and when y is equal to α_0, the general $2n$-point current correlation function vanishes,

$$T^{[1]}_{\mu_1 \cdots \mu_{2n}}(q_1, \ldots, q_{2n}; 0, \alpha_0) = 0 . \tag{74}$$

Finally, let us define a modified two-point function

$$\pi_{2n}^{[1]}(q^2; m, y) \tag{75}$$

by the procedure of linking $2n-2=2(n-1)$ external vertices of the general $2n$-point function with $n-1$ free photon propagators and integrating over the four-momenta carried by these propagators, thus leaving a vacuum-polarization-like tensor of second rank. Because we have enforced current conservation [Eq. (73)] and because there are no photon self-energy insertions, this second-rank tensor has only an over-all logarithmic divergence which can be made finite by a single subtraction. A simple way to perform the subtraction is to use the usual Pauli-Villars procedure of taking the difference of Eq. (75) for two distinct values of the fermion mass, giving the finite expression

$$\begin{aligned}
[\pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y)](-q^2 g_{\mu\nu} + q_\mu q_\nu) = \int \frac{d^4 q_1}{(2\pi)^4} \cdots \frac{d^4 q_{n-1}}{(2\pi)^4} \left(-\frac{i g^{\mu_1\mu_2}}{q_1^{\,2}}\right) \cdots \left(-\frac{i g^{\mu_{2n-3}\mu_{2n-2}}}{q_{n-1}^{\,2}}\right) \\
\times [T^{[1]}_{\mu_1\mu_2 \cdots \mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; m, y) \\
- T^{[1]}_{\mu_1\mu_2 \cdots \mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; m', y)] .
\end{aligned} \tag{76}$$

For the sake of compactness in writing the internal photon propagators, we have restricted ourselves to the Feynman gauge, a convention to which we will adhere henceforth.

Our next step is to establish the following fundamental identity[23] relating the modified two-point function $\pi_{2n}^{[1]}$ to a derivative of the photon proper self-energy part $\pi_c^{[1]}$,

$$2^{n-1} \frac{d^{n-1}}{dy^{n-1}} [\pi_c^{[1]}(q^2; m, y) - \pi_c^{[1]}(q^2; m', y)] = \pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y) . \tag{77}$$

To prove Eq. (77), we develop the right-hand side and the bracket on the left-hand side in power series expansions in y,

$$\pi_c^{[1]}(q^2; m, y) - \pi_c^{[1]}(q^2; m', y) = \sum_{j=0}^{\infty} y^j \pi_{c,j}^{[1]}(q^2; m, m') , \tag{78}$$

$$\pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y) = \sum_{j=0}^{\infty} y^j \pi_{2n,j}^{[1]}(q^2; m, m') ,$$

so that Eq. (77) asserts that

$$2^{n-1} \frac{(j+n-1)!}{j!} \pi_{c, j+n-1}^{[1]}(q^2; m, m') = \pi_{2n, j}^{[1]}(q^2; m, m') . \tag{79}$$

To verify Eq. (79), we proceed in two steps: First, we show that the functions on the left- and right-hand side are the same, apart from a multiplicative constant, and then we give a simple combinatoric argument to show that this constant is in fact $2^{n-1}(j+n-1)!/j!$.

To prove the first assertion, we refer to Fig. 1 defining $\pi_c^{[1]}$ as a power series in y. We see that $\pi_{c,j+n-1}^{[1]}(q^2; m, m')$ is just the sum of all distinct single fermion loop vacuum polarization contributions containing exactly $j+n-1$ internal virtual photons (with the logarithmic divergence eliminated by taking the difference of expressions with fermion masses m and m'). Next, we refer to Fig. 5 and Eq. (76) which respectively define $T_{2n}^{[1]}$ and $\pi_{2n}^{[1]}$. Since the y dependence of $\pi_{2n}^{[1]}$ comes entirely from $T_{2n}^{[1]}$, we see that $\pi_{2n,j}^{[1]}$ contains j internal virtual photons (the ones which appear in the y^j term of $T_{2n}^{[1]}$) plus the $n-1$ additional virtual photons inserted by the definition of Eq. (76), or a total of $j+n-1$ in all. Thus $\pi_{2n,j}^{[1]}(q^2; m, m')$ is also a sum of (mass differenced) single fermion loop vacuum polarization diagrams containing exactly $j+n-1$ internal virtual photons. Furthermore, it is readily seen that all of the relevant diagrams appear in the sum with equal weight because $T_{2n}^{[1]}$ is completely symmetric in the variables of the $2n$ external photons. Hence $\pi_{2n,j}^{[1]}$ must be a multiple of $\pi_{c,j+n-1}^{[1]}$, the constant of proportionality K reflecting the fact that in obtaining the two-point function by linking $2n-2$ external vertices of the $2n$-point function, there

will be multiple counting and each relevant diagram of the two-point function will appear many times.

To complete the derivation of Eq. (79) we must calculate the proportionality constant. This is easily done by noting that

$$K = N\{\pi_{2n,j}^{[1]}\}/N\{\pi_{c,j+n-1}^{[1]}\}, \tag{80}$$

the numerator and denominator in Eq. (80) being the *total* number of distinct Feynman graphs appearing in $\pi_{2n,j}^{[1]}$ and in $\pi_{c,j+n-1}^{[1]}$, respectively. Let us define $N_{2n,j}$ to be the total number of distinct Feynman graphs with j internal virtual photons which contribute to the single fermion loop $2n$-point function. Then from the definitions given above we clearly have

$$\begin{aligned} N\{\pi_{2n,j}^{[1]}\} &= N_{2n,j}, \\ N\{\pi_{c,j+n-1}^{[1]}\} &= N_{2,j+n-1}. \end{aligned} \tag{81}$$

The combinatorics of calculating $N_{2n,j}$ goes as follows. We hold one external vertex fixed on the fermion loop to define a starting point. There are then $(2n+2j-1)!$ diagrams obtained by permuting the remaining $2n-1$ external vertices and the $2j$ vertices which terminate internal photon lines. However, diagrams obtained by permuting any of the j internal photon lines, or interchanging the ends of any of these lines, are identical, and so we must divide by a factor of $2^j j!$ to get the number of distinct Feynman diagrams. Thus we get[24]

$$N_{2n,j} = \frac{(2n+2j-1)!}{2^j j!}, \tag{82}$$

and hence

$$\begin{aligned} K &= \frac{(2n+2j-1)!}{2^j j!} \Big/ \frac{[2+2(j+n-1)-1]!}{2^{j+n-1}(j+n-1)!} \\ &= 2^{n-1}(j+n-1)!/j!, \end{aligned} \tag{83}$$

completing the proof of Eq. (77).

We now have all the apparatus needed to show the existence of an essential singularity. Let us take the limit $m, m' \to 0$ in Eq. (77), with m/m' and q^2 fixed and with $y = \alpha_0$. The left-hand side can be evaluated from the asymptotic expression in Eq. (70), giving

$$2^{n-1}\frac{d^{n-1}}{dy^{n-1}}F^{[1]}(y)\bigg|_{y=\alpha_0} \ln(m'^2/m^2). \tag{84}$$

To evaluate the limit of the right-hand side, we refer to the definition of

$$\pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y)$$

given in Eq. (76). We would like to be able to interchange the subtraction in the square bracket on the right-hand side with the integrations, giving

$$[\pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y)](-q^2 g_{\mu\nu} + q_\mu q_\nu) = I_m - I_{m'}, \tag{85}$$

with

$$I_m = \int \frac{d^4q_1}{(2\pi)^4}\cdots\frac{d^4q_{n-1}}{(2\pi)^4}\left(-\frac{ig^{\mu_1\mu_2}}{q_1^{\,2}}\right)\cdots\left(-\frac{ig^{\mu_{2n-3}\mu_{2n-2}}}{q_{n-1}^{\,2}}\right)T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; m, y). \tag{86}$$

For general values of y, this interchange is not allowed, because I_m is a logarithmically divergent integral of the general type

$$\int_0^\infty \frac{d\rho}{\rho + m^2} \tag{87}$$

and hence the right-hand side of Eq. (85) is an ambiguous expression of the form $\infty - \infty$. When $y = \alpha_0$, however, the situation is different, because Eq. (74) tells us that

$$T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; 0, \alpha_0) = 0 \tag{88}$$

and consequently

$$T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; m, \alpha_0) \tag{89}$$

is proportional to m^2.[25] As a result, the convergence of Eq. (86) is improved by two powers of momentum over what it is for general values of y, and hence when $y = \alpha_0$, I_m becomes a convergent integral of the type[26]

$$\int_0^\infty \frac{cm^2 d\rho}{(\rho + m^2)(\rho + cm^2)}. \tag{90}$$

The interchange in Eq. (85) is now legal,[27] and taking the limit $m, m' \to 0$ gives

$$\begin{aligned} \lim_{\substack{m, m' \to 0 \\ m/m' \text{ fixed}}} & [\pi_{2n}^{[1]}(q^2; m, y) - \pi_{2n}^{[1]}(q^2; m', y)](-q^2 g_{\mu\nu} + q_\mu q_\nu) \\ &= \lim_{m\to 0} I_m - \lim_{m'\to 0} I_{m'} = 0. \end{aligned} \tag{91}$$

Substituting Eqs. (91) and (84) into Eq. (77) we get, finally, the fundamental result

$$\frac{d^{n-1}}{dy^{n-1}} F^{[1]}(y)\bigg|_{y=\alpha_0} = 0\,, \quad n \geq 2\,. \tag{92}$$

It is important to note that Eq. (88) does *not* imply the stronger result

$$\lim_{m\to 0} I_m = 0\,, \tag{93}$$

as is readily seen by taking the $m\to 0$ limit of the specific example in Eq. (90),

$$\lim_{m\to 0} \int_0^\infty \frac{cm^2 d\rho}{(\rho+m^2)(\rho+cm^2)} = \frac{c\ln c}{c-1}\,. \tag{94}$$

Thus, our argument gives us no information about $G^{[1]}(y)$, the fermion-mass independent part of the asymptotic expression for $\pi_c^{[1]}(q^2; m, y)$ given in Eq. (70).

To summarize, we have learned that the function $F^{[1]}(y)$ *and all of its derivatives* are zero at the point $y=\alpha_0$, where α_0 is the zero of the Gell-Mann–Low function $\psi(y)$. In other words, $F^{[1]}$ vanishes with an essential singularity at α_0. It is clear that a similar argument can be applied to the general $2n$-point function $T_{2n}^{[1]}$, by using an identity, analogous to Eq. (77), which relates the derivative $d^m T_{2n}^{[1]}/dy^m$ to an integral over the $(2n+2m)$-point function $T_{2n+2m}^{[1]}$. Thus we additionally learn that when the fermion mass m is zero, the single fermion loop $2n$-point function and all of its y derivatives also vanish at α_0. This fact, together with our result for $F^{[1]}$, gives us information about all of the loop diagrams appearing in the modified skeleton expansion for $f(y)$, from which we learn that f also has an infinite order zero at α_0. Finally, referring to Eq. (61), we conclude that the Gell-Mann–Low function $\psi(y)$ vanishes with an essential singularity at $y=\alpha_0$.[28] In Fig. 6 we summarize the complete chain of reasoning which we have used. Clearly, our conclusion shows that the customary assumption, that α_0 is a simple zero and a point of regularity of ψ, is in fact incorrect.

B. Asymptotic Behavior of h

As we have seen in Sec. II B, the customary assumption about the zero of ψ implies that the nonasymptotic piece h of the photon proper self-energy vanishes with power law behavior,

$$h \sim x^{\psi'(\alpha_0)}\,, \tag{95}$$

as $x=-q^2/m^2$ becomes infinite. Now that we know that ψ actually vanishes with an essential singularity, and not with a simple zero, we must reexamine the reasoning leading to Eq. (95). We give first a general, qualitative argument to show how Eq. (95) must be modified. Let us use the Gell-Mann–Low equation in the form of Eqs. (50) and (51),

$$\ln x = \Phi[\alpha d_c^\infty(x,\alpha)] - \Phi[q(\alpha)]\,,$$

$$\Phi[u] = \int_{c'}^{u} \frac{dz}{\psi(z)}\,. \tag{96}$$

If ψ has a zero at $z=\alpha_0$, then $\Phi[u]$ becomes infinite at $u=\alpha_0$, and hence the large-x behavior of αd_c^∞ is governed by the behavior of Φ in the vicinity of α_0. Now if $\psi(z)$ vanishes more rapidly than $\alpha_0 - z$ as $z\to\alpha_0$, then $v=\Phi[u]$ will become infinite faster than $\ln(\alpha_0-u)$ as $u\to\alpha_0$. This implies that $\Phi^{-1}[v]-\alpha_0$ is a function which is weaker than an exponential as $v\to\infty$, or equivalently,

$$\alpha d_c^\infty(x,\alpha) - \alpha_0 = \Phi^{-1}[\ln x] - \alpha_0 \tag{97}$$

is weaker than a power law as $x\to\infty$. So we obtain the qualitative conclusion that if ψ vanishes more rapidly than with a simple zero as $z\to\alpha_0$, $h(x,\alpha)$ will decrease more slowly than a power law as $x\to\infty$.

To obtain more specifically the connection between the functional form of ψ near α_0 and that of h near $x=\infty$, we resort to the study of exactly integrable examples. As in the discussion of Eqs. (32)–(35), in constructing these examples we can ignore the fact that $\psi(z)$ vanishes at $z=0$, since this region is not relevant to the asymptotic behavior of h. As our first illustration, we consider the case where ψ vanishes with a zero of finite order higher than the first. Substituting

$$\psi(z) = A(\alpha_0 - z)^{1+\epsilon} \tag{98}$$

into the Gell-Mann–Low equation [Eq. (26)] and integrating, we get

$$\alpha d_c^\infty - \alpha_0 = \frac{q(\alpha)-\alpha_0}{\{1+A\epsilon[\alpha_0-q(\alpha)]^\epsilon \ln x\}^{1/\epsilon}} \underset{x\to\infty}{\sim} (\ln x)^{-1/\epsilon}\,, \tag{99}$$

$$\psi(\alpha_0)=0 \Rightarrow \begin{array}{l}\text{General } 2n\text{-point}\\ \text{function} (n\geq 2)\\ \text{vanishes at } y=\alpha_0,\\ m=0\end{array} \Rightarrow \begin{array}{c} F^{[1]}(\alpha_0)=0 \\ T^{[1]}_{2n}\Big|_{\substack{y=\alpha_0\\ m=0}}=0\end{array} \Rightarrow \begin{array}{c} F^{[1]}(\alpha_0)=0^\infty \\ T^{[1]}_{2n}\Big|_{\substack{y=\alpha_0\\ m=0}}=0^\infty\end{array} \Rightarrow f(\alpha_0)=0^\infty \Rightarrow \psi(\alpha_0)=0^\infty$$

FIG. 6. Chain of reasoning which summarizes the discussion of Secs. II D1 and III A. The abbreviation 0^∞ denotes a zero of infinite order (i.e., an essential singularity) in the y variable.

which, as expected, falls off more slowly than any power of x in the limit $x\to\infty$. As our second illustration, we study the case where ψ vanishes with an essential singularity of the form

$$\psi(z)\sim e^{-A/(\alpha_0-z)^p}\,. \tag{100}$$

To get an exactly integrable expression we multiply Eq. (100) by a power of α_0-z, giving

$$\psi(z)=\frac{(\alpha_0-z)^{p+1}}{ABp}\,e^{-A/(\alpha_0-z)^p}\,. \tag{101}$$

Substituting Eq. (101) into Eq. (26) and doing the z integration, we get

$$\alpha d_c^\infty-\alpha_0=\frac{-A^{1/p}}{(\ln[B^{-1}\ln x+\exp\{A/[\alpha_0-q(\alpha)]^p\}])^{1/p}}$$

$$\underset{x\to\infty}{\sim}(\ln\ln x)^{-1/p}\,. \tag{102}$$

We see that when ψ vanishes with an essential singularity at α_0, the asymptotic vanishing of h in the limit of large x is very slow indeed. In Table I we summarize the connection between the type of zero of ψ and the asymptotic behavior of h that we have inferred from our examples.[29]

C. Consequences of the Slow Decrease of h

In both the justification of the JBW form of the eigenvalue condition [Eq. (59) and the discussion preceding it in Sec. II D] and the derivation of the scaling form of the asymptotic electron propagator [Appendix B] we assume that α_0 is a simple zero, and a point of regularity, of the Gell-Mann–Low function ψ, and that h decreases asymptotically with power-law behavior. Now that we have seen that these assumptions are false, we must reexamine our treatment of the eigenvalue condition and of the asymptotic behavior of the electron propagator, to study the consequences of the essential singularity which we have found in ψ and of the concomitant very slow asymptotic decrease of h. For the sake of definiteness, we will assume behavior as in Eqs. (101) and (102) with $p=1$, so that h decreases asymptotically as

$$h\sim\frac{1}{\ln\ln(-q^2/m^2)}\,. \tag{103}$$

TABLE I. Connection between behavior of $\psi(z)$ near $z=\alpha_0$ and behavior of $h(x,\alpha)$ near $x=\infty$.

Behavior of ψ near α_0	Asymptotic behavior of h
$\psi'(\alpha_0-z)$	$x^{\psi'}$
$(\alpha_0-z)^{1+\epsilon}$	$(\ln x)^{-1/\epsilon}$
$e^{-A/(\alpha_0-z)^p}$	$(\ln\ln x)^{-1/p}$

This restriction, while convenient to make, is not crucial to the discussion which follows.

Let us first reconsider the eigenvalue condition, picking up our discussion of Sec. II D at the point where we established that logarithmic behavior of a graph contributing to π_c can only be associated with the over-all integration involving all lines in the graph. As we noted, each internal photon line in the over-all integration contributes two parts, a part proportional to α_0 and a part proportional to h. Let us separately group together all contributions to π_c involving no factors of h, all contributions involving exactly one factor of h, all those involving exactly two factors of h, etc., as indicated in Fig. 7. The shaded blobs in Fig. 7, to which the insertions of h are attached, are two-point, four-point, six-point, etc. functions calculated with all internal photons described by free propagators coupling with the asymptotic coupling strength α_0. The piece with no factors of h is just the one retained in our earlier discussion, which, as we have seen, makes the contribution

$$g(\alpha_0)+f(\alpha_0)\ln(-q^2/m^2) \tag{104}$$

to the asymptotic behavior of π_c. Heuristically speaking, the logarithm in Eq. (104) can be thought of as arising from the integral

$$\int_{m^2}^{-q^2}\frac{d\rho}{\rho}\,; \tag{105}$$

in this language, the leading asymptotic behavior of the piece of π_c containing n factors of h corresponds to the integral

$$\int_{m^2}^{-q^2}\frac{d\rho}{\rho}\,h(\rho/m^2,\alpha)^n\,. \tag{106}$$

When h vanishes as a power of ρ for large ρ, the integral in Eq. (106) converges at the upper limit as $-q^2/m^2\to\infty$. Asymptotic finiteness of π_c then only requires the vanishing of the coefficient of the integral in Eq. (105), giving the JBW condition $f(\alpha_0)=0$. When h vanishes much more slowly than a power law, as in Eq. (103), the situation is radically changed.[30] The integral in Eq. (106) is now

$$\int_{m^2}^{-q^2}\frac{d\rho}{\rho[\ln\ln(\rho/m^2)]^n}\,, \tag{107}$$

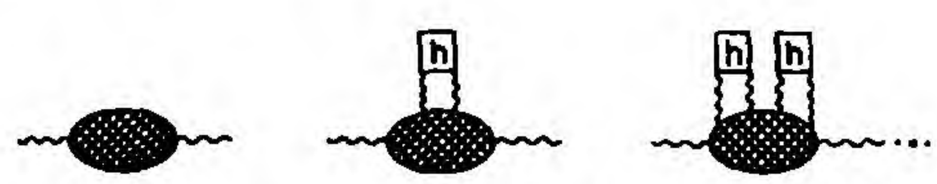


FIG. 7. Grouping of π_c into contributions involving no factor h, exactly one factor h, exactly two factors h, etc. The shaded blobs denote two-point, four-point, six-point, etc. functions calculated with all internal photons described by free propagators coupling with the asymptotic coupling strength α_0.

which for all n is divergent at the upper limit as $-q^2/m^2 \to \infty$. Thus, asymptotic finiteness of π_c requires now that an infinite number of conditions be satisfied: in addition to the coefficient of Eq. (105) vanishing, the coefficient of the contribution represented heuristically by Eq. (107) must vanish for all n. It is remarkable that when α_0 is chosen to be the root of $f(\alpha_0)=0$, this infinity of conditions is in fact satisfied. The reason is the argument based on the Federbush-Johnson theorem given in Eqs. (62)–(63) of Sec. II D 1, which shows that when $f(\alpha_0)=0$ and the fermion mass m is zero, the general $2n$-point current correlation function vanishes for $n \geq 2$. Hence when α_0 satisfies $f(\alpha_0)=0$, each shaded blob in Fig. 7 is proportional to m^2 and therefore contributes a convergence factor m^2/ρ to the integral in Eq. (107). The integral then becomes[25]

$$\int_{m^2}^{-q^2} \frac{d\rho\, m^2}{\rho^2[\ln\ln(\rho/m^2)]^n}, \tag{108}$$

which is asymptotically finite as $-q^2/m^2 \to \infty$. The asymptotically divergent integral in Eq. (107) of course reappears when α_0 is chosen to have any value other than the root of $f(\alpha_0)=0$. We conclude, then, that Eq. (103) still permits one to deduce the JBW eigenvalue condition $f(\alpha_0)=0$, but only by a more involved mechanism than is required in the case of a power law vanishing of h.

Let us next examine the implications of the essential singularity at α_0 and of Eq. (103) for the argument leading to the scaling form for the asymptotic electron propagator. As we have noted, the approach used to derive the scaling form in Appendix B depends very specifically on the assumptions of regularity of the theory in the vicinity of α_0 and power law vanishing of h. To deal with the situation where α_0 is a point of essential singularity, we give an alternative approach, based on reasoning similar to that which we have just used in our discussion of the eigenvalue condition. Let us consider the *unrenormalized* electron propagator $S_F'(p)^{-1}$ in the limit in which $-p^2$ and the cutoff Λ^2 are both becoming infinite relative to the fermion mass m^2. To study this, we collect together all contributions to the electron proper self-energy involving no factors of h, involving exactly one factor of h, exactly two factors of h, etc., as shown in Fig. 8. As before, the shaded blobs are calculated with all internal photons described by free propagators coupling with the asymptotic coupling strength α_0. The piece with no factors of h is just the unrenormalized electron proper self-energy in the JBW model. A straightforward analysis[31] using the methods of Ref. 17 shows that if this piece alone is retained, the unrenormalized asymptotic electron propagator has the scaling form

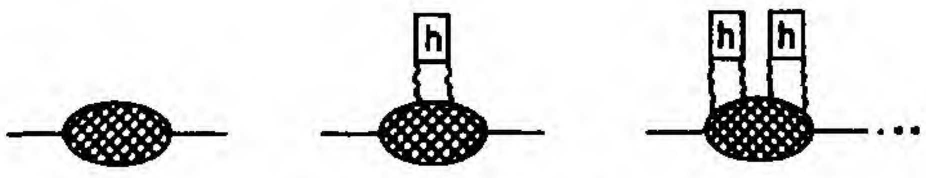


FIG. 8. Grouping of the electron proper self-energy into contributions involving no factors h, exactly one factor h, exactly two factors h, etc. The shaded blobs are calculated with all internal photons described by free propagators coupling with the asymptotic coupling strength α_0.

$$S_F'(p)^{-1} \underset{\substack{-p^2/m^2\to\infty \\ \Lambda^2/m^2\to\infty}}{\sim} F_1(\alpha_1)\left(-\frac{p^2}{\Lambda^2}\right)^{\gamma(\alpha_1)/2} \times\left[\not{p} - m_0 F_2(\alpha_1)\left(-\frac{p^2}{\Lambda^2}\right)^{-\delta(\alpha_1)/2}\right]. \tag{109}$$

Together with the fact that S_F' and the scalar vertex Γ_S are multiplicatively renormalizable, Eq. (109) implies[31] the results of Eq. (66) for both the renormalized electron propagator and the renormalization constants Z_2 and m_0, with the modification, already noted in Sec. II D, that the constants C_1 and C_2 in Eq. (66) become dependent on nonasymptotic quantities. We must now examine whether the asymptotic expression of Eq. (109) is modified by the pieces containing one or more factors of h. To this end, it is useful to note that the powers in Eq. (109) arise in perturbation theory from infinite sums of logarithms,

$$\left(-\frac{p^2}{\Lambda^2}\right)^{\gamma(\alpha_1)/2} = \sum_{n=0}^{\infty} \frac{[\frac{1}{2}\gamma(\alpha_1)\ln(-p^2/\Lambda^2)]^n}{n!}, \tag{110}$$

and heuristically, the logarithms can be thought of as arising from integrals of the form

$$\int_{-p^2}^{\Lambda^2} \frac{d\rho}{\rho}. \tag{111}$$

In this language, the piece of the electron proper self-energy containing n factors of h will involve integrals of the form

$$\int_{-p^2}^{\Lambda^2} \frac{d\rho}{\rho} h(\rho/m^2, \alpha)^n. \tag{112}$$

If h vanishes as a power of ρ for large ρ, the integral in Eq. (112) vanishes as $-p^2/m^2$, $\Lambda^2/m^2 \to \infty$, and the scaling form of Eq. (109) is unmodified.[30] On the other hand, if h vanishes as in Eq. (103), then Eq. (112) becomes

$$\int_{-p^2}^{\Lambda^2} \frac{d\rho}{\rho[\ln\ln(\rho/m^2)]^n}, \tag{113}$$

which does *not* vanish[30] in the limit of asymptotic $-p^2$, Λ^2 and which could therefore give rise to corrections to Eq. (109). We again can salvage the situation if we can use the Federbush-Johnson theorem to argue that the Compton-like shaded blobs in Fig. 8 vanish when $f(\alpha_0)=0$ and the fermion mass m is zero. However, this involves an extension of the Federbush-Johnson theorem outside the charge-zero sector, which is the only place where a satisfactory proof in the case of electrodynamics has been given.[8,32] If such an extension is allowed, we gain a convergence factor m^2/ρ in Eq. (113), giving

$$\int_{-p^2}^{\Lambda^2} \frac{d\rho m^2}{\rho^2[\ln\ln(\rho/m^2)]^n}, \tag{114}$$

which vanishes as $-p^2/m^2$, $\Lambda^2/m^2 \to \infty$.

We conclude, then, that the JBW eigenvalue condition and, possibly, the scaling form for the asymptotic electron propagator remain valid in the presence of the essential singularity, but only by virtue of an additional infinity of conditions being satisfied simultaneously. This, of course, poses troublesome questions of convergence (basically, is $0\times\infty$ effectively 0 in these problems?) which we have not attempted to settle.

IV. LOOPWISE SUMMATION AND AN EIGENVALUE CONDITION FOR α

Up to this point we have consistently employed the "vacuum-polarization-insertion-wise" summation scheme, both in our review of the JBW results in Sec. II D and in our deduction of the presence of an essential singularity in the preceding section. As we have seen, this scheme leads to a one-parameter family of asymptotically finite solutions, in which the asymptotic coupling α_0 is determined to be the zero y_0 of the Gell-Mann–Low function $\psi(y)$ [and simultaneously a zero of the simpler functions $f(y)$ and $F^{[1]}(y)$], while the physical coupling α is a free parameter, restricted only by the condition $\alpha < \alpha_0 = y_0$ following from spectral function positivity [see Eq. (129) below.] The usual assumption is that this one-parameter family represents the most general type of asymptotically finite solution which can occur. In the present section, we show that the presence of a simultaneous zero in all of the single fermion-loop diagrams *makes possible one additional asymptotically finite solution*, which has the very appealing feature that the physical coupling α is fixed to be y_0. Our procedure is not strictly deductive, in that we continue to accept the *results* concerning properties of the single fermion-loop diagrams which were found in Sec. III A, while dropping the identification of α_0 with y_0 which was made there. We will also introduce a new order of summing the perturbation series, involving "loopwise" rather than "vacuum-polarization-insertion-wise" summation. Specifically, we make the following two assumptions:

(1) The function $F^{[1]}(y)$ defined by Fig. 1 and the $2n$-point current correlation function with zero fermion mass, $T^{[1]}_{\mu_1\cdots\mu_{2n}}(q_1,\ldots,q_{2n};m=0,y)$, vanish simultaneously at $y=y_0$. As we have seen in Sec. III A, the simultaneous vanishing of $F^{[1]}$ and $T^{[1]}_{2n}$ implies that they vanish with a zero of infinite order.

(2) The photon proper self-energy can be correctly obtained by "loopwise" summation. That is, we assume convergence of the sum

$$\pi_c = \sum_{n=1}^{\infty} \pi_c^{[n]}, \tag{115}$$

where $\pi_c^{[n]}$ is the contribution to the photon proper self-energy containing exactly n closed fermion loops. The burden of the present section will be to show that *these two assumptions imply asymptotic finiteness of the photon propagator when the physical fine structure constant is chosen to have the value $\alpha=y_0$. Furthermore, we will show that for this particular value of α the function $\beta(\alpha)$ appearing in the Callan-Symanzik equation vanishes (when summed loopwise) and so the theory has type-1 asymptotic behavior.*

To proceed, we introduce some additional definitions. Let $\beta^{[n]}(\alpha)$ be the contribution to $\beta(\alpha)$ with exactly n closed fermion loops, and let $\pi_c^{[n,r]}$ be the part of $\pi_c^{[n]}$ in which exactly r closed fermion loops remain when all internal photon self-energy parts are shrunk down to points [see Fig. 9.] In terms of these definitions, we can write

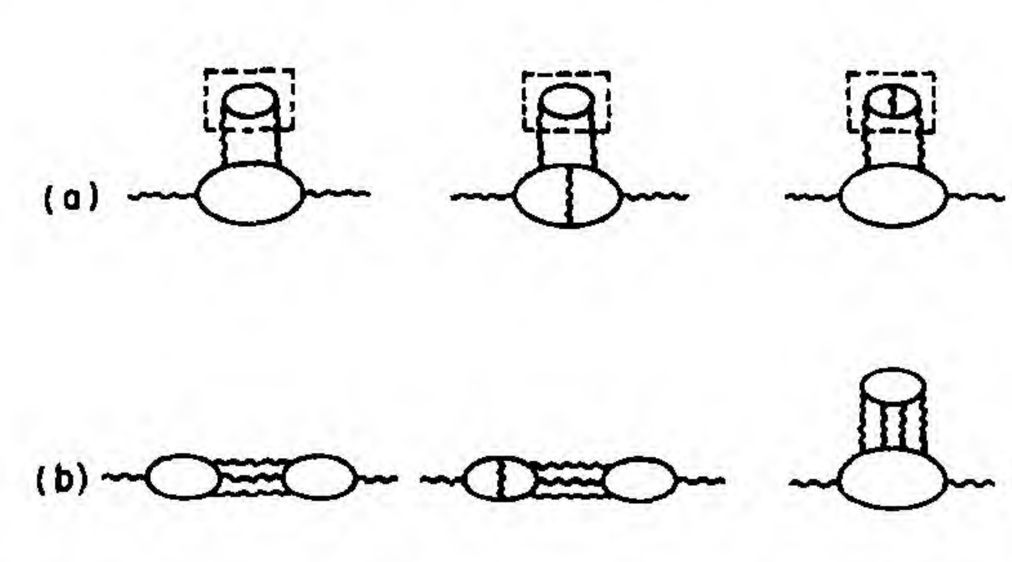


FIG. 9. (a) Typical diagrams contributing to $\pi_c^{[2,1]}$, the part of the two-fermion-loop photon proper self-energy which contains only one fermion loop after the internal photon self-energy part (enclosed by dashed lines) is shrunk down to a point. (b) Typical diagrams contributing to $\pi_c^{[2,2]}$, the part of the two-fermion-loop photon proper self-energy which still contains two fermion loops after shrinking away the internal photon self-energy parts.

$$\beta(\alpha)=\sum_{n=1}^{\infty}\beta^{[n]}(\alpha),$$
$$\pi_c^{[n]}=\sum_{r=1}^{n}\pi_c^{[n,r]}. \tag{116}$$

We now begin our argument by considering the case $n=1$. Because we are dealing with the renormalized theory, the coupling constant which appears is the physical fine structure constant α, and so (using our earlier notation) we must study the asymptotic behavior of $\pi_c^{[1]}(q^2;m,\alpha)$. Referring back to Eq. (70), we see that for asymptotic $-q^2/m^2$ we have

$$\pi_c^{[1]}(q^2;m,\alpha)=G^{[1]}(\alpha)+F^{[1]}(\alpha)\ln(-q^2/m^2) + \text{vanishing terms}; \tag{117}$$

hence choosing $\alpha=y_0$ guarantees the asymptotic finiteness of $\pi_c^{[1]}$. Next we consider the case $n=2$, for which we can write

$$\pi_c^{[2]}=\pi_c^{[2,1]}+\pi_c^{[2,2]}, \tag{118}$$

with the two terms in Eq. (118) corresponding respectively to the diagrams in Figs. 9(a) and 9(b). Because the single fermion-loop vacuum-polarization insertion has already been shown to be asymptotically finite, we can use the argument which was employed above in getting Eq. (58) to show that $\pi_c^{[2,1]}$, as well as $\pi_c^{[2,2]}$, grows asymptotically at worst as a single power of $\ln(-q^2/m^2)$,

$$\pi_c^{[2,1]}(q^2;m,\alpha)=G^{[2,1]}(\alpha)+F^{[2,1]}(\alpha)\ln(-q^2/m^2) + \text{vanishing terms},$$
$$\pi_c^{[2,2]}(q^2;m,\alpha)=G^{[2,2]}(\alpha)+F^{[2,2]}(\alpha)\ln(-q^2/m^2) + \text{vanishing terms}. \tag{119}$$

Furthermore, the same argument tells us that the potential logarithm is associated with the subintegrations involving all lines in $\pi_c^{[2,2]}$, and all lines in $\pi_c^{[2,1]}$ which remain after the internal photon self-energy part has been shrunk down to a point. Clearly, these subintegrations always involve at least one single fermion loop $2j$-point function (with $j\geq 2$) which, we have assumed, vanishes when $\alpha=y_0$ and the fermion mass m is zero. As a result, the potentially dangerous subintegrations are really two powers of momentum more convergent than indicated by naive power counting [cf. Eq. (90)] and hence cannot actually lead to logarithmic asymptotic behavior. So we learn that when $\alpha=y_0$, we have $F^{[2,1]}(\alpha)=F^{[2,2]}(\alpha)=0$, and therefore $\pi_c^{[2]}$ is asymptotically finite. Note that the argument which we have just given does not determine the actual limiting values of $\pi_c^{[1]}$ or $\pi_c^{[2]}$, i.e., we learn nothing about the values of $G^{[1]}(\alpha)$, $G^{[2,1]}(\alpha)$, or $G^{[2,2]}(\alpha)$ at $\alpha=y_0$. This is expected, because the G's depend on the *nonasymptotic* theory (where m cannot be neglected) as a result of the subtraction at $q^2=0$ which renormalizes the photon proper self-energy. Since knowledge of the G's would allow one to calculate α_0 through the formula

$$\sum_{n=1}^{\infty}\sum_{r=1}^{n}G^{[n,r]}(\alpha)=\alpha_0^{-1}-\alpha^{-1}, \tag{120}$$

we see that in our solution with α fixed, α_0 cannot be determined through asymptotic considerations alone.

The next step in the argument is to prove the vanishing of $\beta^{[1]}(\alpha)$ at $\alpha=y_0$. We do this by using the Callan-Symanzik equation in the form given by Eq. (43) which, on substituting Eq. (12) for d_c^{-1}. and dropping the asymptotically vanishing term proportional to $\tilde{\Gamma}_{\gamma\gamma S}$, becomes

$$-\beta(\alpha)+\left[m\frac{\partial}{\partial m}+\beta(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\right]\alpha\pi_c\sim 0. \tag{121}$$

The one-fermion-loop part of this equation is

$$-\beta^{[1]}(\alpha)+m\frac{\partial}{\partial m}\alpha\pi_c^{[1]}\sim 0. \tag{122}$$

Substituting Eq. (70) for $\pi_c^{[1]}$ this becomes

$$\beta^{[1]}(\alpha)=-2\alpha F^{[1]}(\alpha), \tag{123}$$

from which we immediately learn that $\beta^{[1]}(\alpha)$ vanishes at $\alpha=y_0$.

We now continue the argument inductively. We assume that when $\alpha=y_0$ the pieces $\pi_c^{[1]},\ldots,\pi_c^{[n]}$ of the photon proper self-energy are asymptotically finite, while the pieces $\beta^{[1]},\ldots,\beta^{[n]}$ of the Callan-Symanzik function β are zero. We wish to extend these assertions to the pieces $\pi_c^{[n+1]}$ and $\beta^{[n+1]}$ which contain one more closed fermion loop. We write

$$\pi_c^{[n+1]}=\sum_{r=1}^{n+1}\pi_c^{[n+1,r]}, \tag{124}$$

where, according to our induction hypothesis and the argument preceding Eq. (58), the piece $\pi_c^{[n+1,r]}$ can grow asymptotically at most as a single power of $\ln(-q^2/m^2)$,

$$\pi_c^{[n+1,r]}(q^2;m,\alpha)=G^{[n+1,r]}(\alpha) + F^{[n+1,r]}(\alpha)\ln(-q^2/m^2) + \text{vanishing terms}. \tag{125}$$

Again, the argument leading to Eq. (125) tells us that the potential logarithm is associated with the subintegration involving all lines in $\pi_c^{[n+1,r]}$ which remain after the internal photon self-energy parts have been shrunk away. This subintegration always involves at least one single fermion-loop $2j$-point function ($j\geq 2$) which, when $\alpha=y_0$, improves the ultraviolet convergence of the subintegration

by two powers of momentum and hence prevents a logarithm from actually appearing in Eq. (125). So we conclude that $F^{[n+1,r]}(\alpha)=0$ when $\alpha=y_0$, $r=1,\ldots,n+1$, and hence $\pi_c^{[n+1]}$ is asymptotically finite. To prove the vanishing of $\beta^{[n+1]}(\alpha)$, we consider the part of Eq. (120) involving exactly $n+1$ closed fermion loops,

$$-\beta^{[n+1]}(\alpha)+m\frac{\partial}{\partial m}\alpha\pi_c^{[n+1]} + \sum_{r=1}^{n}\beta^{[r]}(\alpha)\left(\alpha\frac{\partial}{\partial\alpha}-1\right)\alpha\pi_c^{[n+1-r]}\sim 0\,. \tag{126}$$

Using the induction hypothesis on β this equation simplifies, when $\alpha=y_0$, to

$$-\beta^{[n+1]}(\alpha)+m\frac{\partial}{\partial m}\alpha\pi_c^{[n+1]}\sim 0\,. \tag{127a}$$

But asymptotic finiteness of $\pi_c^{[n+1]}$ tells us that

$$m\frac{\partial}{\partial m}\alpha\pi_c^{[n+1]}\sim 0\,, \tag{127b}$$

so we learn that $\beta^{[n+1]}(\alpha)=0$ when $\alpha=y_0$, completing the induction.

To summarize, we have learned, for all n, that $\pi_c^{[n]}$ is asymptotically finite and that $\beta^{[n]}$ vanishes when $\alpha=y_0$. Invoking now our assumption of convergence of the "loopwise" summations in Eq. (115) and Eq. (116), we learn that when $\alpha=y_0$, the total photon proper self-energy π_c is asymptotically finite, and the total Callan-Symanzik function $\beta(\alpha)$ vanishes. The vanishing of the Callan-Symanzik function means that our solution with $\alpha=y_0$ has type-1 asymptotic behavior. According to the discussion of Appendix B, the asymptotic electron propagator must then have the simple scaling form of Eq. (66) (with $\alpha_1=\alpha$), leading, as we have noted, to the possibility of a finite m_0 and Z_2.

In conclusion, we briefly discuss the relation of the asymptotically finite solution which we have just found to the "vacuum-polarization-insertion-wise" summation methods used earlier. As we have seen, in our "loopwise" solution α is determined by the condition $F^{[1]}(\alpha)=0$, with the asymptotic coupling α_0 determined from α by the functions $G^{[n,r]}(\alpha)$ according to Eq. (120). *A priori*, we can say nothing about the value of α_0 except that positivity of the spectral function $w(\rho,\alpha)$ appearing in the Källén-Lehmann representation[33] for the photon propagator,

$$d_c(-q^2/m^2,\alpha)=1+q^2\int_0^\infty w(\rho/m^2,\alpha)\frac{d(\rho/m^2)}{q^2-\rho-i\epsilon}\,, \tag{128}$$

implies the sum rule[34]

$$\alpha_0=\alpha+\int_0^\infty \alpha w(\rho/m^2,\alpha)d(\rho/m^2)\,, \tag{129}$$

and hence $\alpha_0>\alpha$. This inequality raises an apparent paradox when we turn to the "vacuum-polarization-insertion-wise" summation scheme, which if applicable would imply that α_0 obeys the same eigenvalue condition as does α, $F^{[1]}(\alpha_0)=0$. The paradox is resolved, however, when we note that since y_0 is an essential singularity of $F^{[1]}(y)$, the point $\alpha_0>\alpha=y_0$ lies *outside* the radius of convergence of $F^{[1]}(y)$, and so the interchange of limit and sum leading to the eigenvalue condition on α_0 is unjustified. Another way of stating this is obtained by writing down the formal Taylor expansion connecting the eigenvalue conditions for α and α_0,

$$F^{[1]}(\alpha_0)=\sum_{n=0}^{\infty}\frac{(\alpha_0-y_0)^n}{n!}\frac{d^n}{dy^n}F^{[1]}(y)\bigg|_{y=y_0=\alpha}\,. \tag{130}$$

Since $F^{[1]}$ and all its derivatives vanish at y_0, naive application of Eq. (130) tells us that $F^{[1]}(\alpha_0)=0$. This conclusion is of course incorrect, because the Taylor expansion in Eq. (130) attempts the analytic continuation of $F^{[1]}$ outside its region of regularity, and therefore is mathematically meaningless. In other words, because of the essential singularity, we cannot freely rearrange the "loopwise"-summed theory, with $\alpha=y_0$, into a "vacuum-polarization-insertion-wise"-summed theory.

V. DISCUSSION

We have learned that there are two possible ways of having an asymptotically finite electrodynamics. The first is the usual one-parameter family of solutions, in which the asymptotic coupling α_0 is fixed to be y_0 and the physical coupling $\alpha<\alpha_0$ is a free parameter. The second is the unique additional solution found in the preceding section, in which the physical coupling α is fixed to be y_0. *We conjecture that nature in fact chooses this second type of solution, and hence that the fine structure constant may be calculated by determining the location of the infinite order zero y_0 of the function $F^{[1]}(y)$.*[35] [Of course, if the function $F^{[1]}(y)$ does *not* have an infinite-order positive zero, then electrodynamics *cannot* be asymptotically finite.] We can advance two possible reasons why nature may choose the solution which fixes α over the solutions which fix α_0:

(1) The "vacuum-polarization-insertion-wise" summation procedure needed to get the solutions which fix α_0 may be divergent for all nonzero values of α. In other words, electrodynamics may exist only when summed "loopwise," with the spe-

cific choice of physical coupling $\alpha = y_0$.

(2) Both types of solution may be mathematically valid, but there may be stability arguments which tell us that when other interactions (such as weak or gravitational interactions) are switched on, the theory chooses the largest possible value of α, that is $\alpha = y_0$.

We emphasize that we have given no arguments which distinguish which, if either, of these possible reasons is correct.

We conclude the paper by giving a brief, speculative discussion of some further implications of the work of the preceding sections. First, we point out a possible connection of our work with Dyson's[9] old conjecture suggesting singularities in electrodynamics at $\alpha = 0$. Then, we discuss the fact that the conjecture stated at the beginning of this section gives a *species-independent* determination of α, and give an argument based on this which may justify our neglect of strong interaction corrections.

A. Dyson's Conjecture

Dyson has argued that the renormalized perturbation theory of quantum electrodynamics, regarded as a power series in α, cannot have a nonzero radius of convergence. For if it did, the theory would still exist when analytically continued to negative α, which corresponds to a physical situation in which *like* charges, rather than unlike charges, attract. But in the analytically continued theory, the usual vacuum, defined as the state containing no particles, would be unstable. To see this, we note that if we create N electron positron pairs, with N very large, and group the electrons together in one region of space and the positrons together in another separate region, we can create a pathological state in which the negative potential energy of the Coulomb forces exceeds the total rest energy and kinetic energy of the particles. Although this state is separated from the usual vacuum by a high potential barrier (of the order of the rest energy of the $2N$ particles being created), quantum-mechanical tunneling from the vacuum to the pathological state would be allowed, and would lead to an explosive disintegration of the vacuum by spontaneous polarization. This instability means that electrodynamics with negative α cannot be described by well-defined analytic functions; hence the perturbation series of electrodynamics must have zero radius of convergence.

If one assumes, as we do in this paper, that electrodynamics is by itself a complete theory,[36] then physical quantities in electrodynamics are described by well-defined, calculable functions of α when α is positive. According to Dyson's argument however, these functions cannot be continued to negative α, and therefore must have a singularity at $\alpha = 0$. Because the singularity originates in a tunneling phenomenon, and because tunneling amplitudes are typically negative exponentials of a barrier-penetration factor, it is plausible that this singularity should be an essential singularity of the form $e^{-c/\alpha}$.

We can now attempt to make a connection with the results of the preceding two sections. As we recall, we argued there that the single-fermion-loop function $F^{[1]}(\alpha)$ should possess an essential singularity (perhaps of the form $\exp[-c/(y_0 - \alpha)]$, resembling a tunneling amplitude) at the point $\alpha = y_0 > 0$. In establishing a connection with Dyson's work, there appear to be two possibilities. One possibility is that the singularity at y_0 is *not* Dyson's singularity, but that electrodynamics exists for a range of positive α and that $F^{[1]}(\alpha)$ (or perhaps some other function in the theory) has a singularity at $\alpha = 0$ which prevents continuation to negative α. An alternative possibility is that $F^{[1]}(\alpha)$ is regular at $\alpha = 0$, but that the full photon propagator simply does not exist for values of the physical coupling α other than y_0, preventing continuation of the complete theory to negative fine-structure constant. In this case, the singularity of $F^{[1]}$ at y_0 would be a mathematical manifestation of Dyson's argument. In this connection, it is intriguing that the class of single-fermion-loop vacuum-polarization diagrams which we assert to possess an essential singularity are just the simplest diagrams describing the creation of an arbitrarily large number of pairs from the vacuum, and therefore are the simplest diagrams leading to Dyson's pathological state. For, as shown in Fig. 10, the single-fermion-loop diagrams describe the creation of an arbitrary number of pairs from the vacuum, but with only one fermion world line actually present.

B. Species Independence

Up to this point we have assumed the presence of only one species of fermion interacting solely

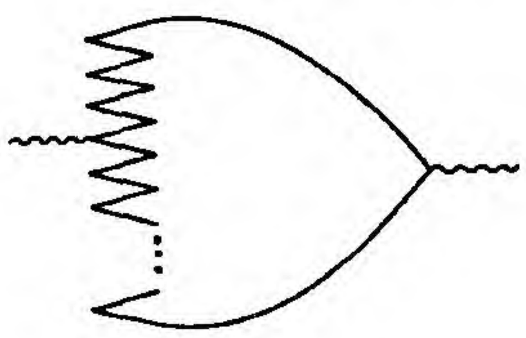


FIG. 10. Ordering in which a single-fermion vacuum-polarization loop diagram describes the creation of an infinite number of pairs from the vacuum. (We have not drawn in any of the internal photons.)

with the electromagnetic field. Let us now consider the more general case in which there are j elementary charged spin-$\frac{1}{2}$ fermion species which, for the moment, we still assume to interact only electromagnetically. Although these fermions may have different masses, the contributions of mass-difference terms to the photon proper self-energy are guaranteed, just by power counting, to be asymptotically finite. Hence to study the effect of having j fermions on the asymptotic behavior of the photon propagator, it suffices to consider the special case in which they all have a common mass m. Then, because each closed fermion loop in the photon proper self-energy appears j times, the piece of π_c containing exactly n closed fermion loops is multiplied by j^n, and so Eq. (115) is modified to read

$$\pi_c = \sum_{n=1}^{\infty} j^n \pi_c^{[n]} . \tag{131}$$

Clearly, because choosing $\alpha = y_0$ makes each of the $\pi_c^{[n]}$ individually asymptotically finite, this choice of coupling makes the total π_c asymptotically finite as well, independent of the species number j. Stated in another way, when j fermion species are present the single fermion loop function determining y_0 is just $j F^{[1]}(y)$, and so the value of y_0 determined is the same as in the one-species case. Thus we reach the important conclusion that *our eigenvalue condition for determining α is independent of the fermion species number*. Whether this species independence is maintained in the presence of elementary charged spin-0 boson fields is not clear. The requirement is obviously that the function $F_B^{[1]}(y)$, defined by summing the single charged boson loop diagrams of Fig. 1 in analogy to our definition of $F^{[1]}(y)$, must vanish with an infinite order zero at the same point y_0 where $F^{[1]}(y)$ vanishes. All that is known about $F^{[1]}(y)$ and $F_B^{[1]}(y)$ at present is the first few terms in their respective power-series expansions,[37]

$$\begin{aligned} -yF^{[1]}(y) &= \frac{2}{3}\left(\frac{y}{2\pi}\right)+\left(\frac{y}{2\pi}\right)^2-\frac{1}{4}\left(\frac{y}{2\pi}\right)^3+\cdots, \\ -yF_B^{[1]}(y) &= \frac{1}{6}\left(\frac{y}{2\pi}\right)+\left(\frac{y}{2\pi}\right)^2+\cdots . \end{aligned} \tag{132}$$

Equation (132) tells us that the functions $F^{[1]}(y)$ and $F_B^{[1]}(y)$ are not identical, but of course says nothing about their behavior when summed to all orders.

Returning, now, to our model with several charged fermion species, let us suppose that some of these fermions have strong interactions mediated by neutral boson exchange (the gluon model).

FIG. 11. Fermion vacuum-polarization loop modified by internal gluon (dashed line) radiative corrections.

Although the bosons do not themselves contribute vacuum-polarization loops, they could modify the fermion vacuum polarization loops when they appear as internal radiative corrections (see Fig. 11.) However, let us now invoke the experimental observation of scaling in deep-inelastic electron scattering, one explanation for which[38] is that the exchanges which mediate the strong interactions are actually much more strongly damped at high four-momentum transfer than is the free boson propagator $(q^2+\mu^2)^{-1}$. If such an explanation proves correct,[39] then vacuum-polarization diagrams with gluon radiative corrections will by themselves be asymptotically finite, and so the presence of strong interactions will not alter our eigenvalue condition for α. Our scheme is clearly incompatible, however, with the presence of fractionally charged fermions such as quarks[40]; all elementary charged fermions must have the same basic electromagnetic coupling $(\pm)\sqrt{\alpha}$.

Note added in proof. R. F. Dashen has pointed out to us that in order y^3 and higher the vacuum polarization structure of charged spin-0 boson electrodynamics will differ from that of the spin-$\frac{1}{2}$ case, as a result of the presence of a boson-boson scattering counterterm in the Lagrangian. Hence the analysis which we have given above for the case of spin-$\frac{1}{2}$ electrodynamics cannot be directly applied to the spin-0 case. The JBW argument for finiteness of the bare mass also fails in spin-0 electrodynamics. [See D. Flamm and P. G. O. Freund, Nuovo Cimento 32, 486 (1964).]

ACKNOWLEDGMENTS

I want to thank R. F. Dashen and W. A. Bardeen for discussions which led to this work, and F. J. Dyson, B. Simon, and S. B. Treiman for a careful critical reading of the manuscript. I am grateful to the following people for helpful conversations, for criticism and/or for reading the manuscript: C. G. Callan, M. L. Goldberger, D. J. Gross, K. Johnson, A. Pais, A. Sirlin, F. Strocchi, A. S. Wightman, K. Wilson, and D. R. Yennie. I wish to thank Angela Gonzales of the National Accelerator Laboratory staff for preparing the figure drawings.

APPENDIX A: PARTIAL SUMMARY OF NOTATION

JBW	Johnson-Baker-Willey
α	physical coupling (fine-structure constant)
α_w	new coupling constant defined by subtraction at w
α_0	asymptotic coupling constant
α_b	canonical or bare coupling constant, related to α by $\alpha_b = Z_3^{-1}\alpha$
α_1	root of $q(\alpha_1)=\alpha_0$, with $q(y)=yd_c^\infty(1,y)$
Z_3	photon wave-function renormalization constant
m	electron physical mass
$\alpha d_c(-q^2/m^2,\alpha)$	renormalized photon propagator; $d_c(-q^2/m^2,\alpha)=[1+\alpha\pi_c(q^2)]^{-1}$
$h(-q^2/m^2,\alpha)$	difference between αd_c and its asymptotic limit α_0
$\alpha d_c^\infty(-q^2/m^2,\alpha)$	"asymptotic part" of the renormalized photon propagator, obtained by dropping in each order of perturbation theory terms which vanish as $-q^2/m^2\to\infty$, but keeping terms in each order which are constant or increase logarithmically
$\psi(y)$	Gell-Mann–Low function
$F^{[1]}(y)$	coefficient of logarithmically divergent part of the sum of single-fermion-loop vacuum polarization diagrams
y_0	point where $F^{[1]}(y)$ has an infinite-order zero (essential singularity)
Λ	ultraviolet cutoff
μ, μ_0	physical photon mass (infrared cutoff), bare photon mass
$D_F^0(q)_{\mu\nu}$	bare photon propagator
ξ	gauge parameter (coefficient of longitudinal part of photon propagator)
$\pi(q^2)_{\mu\nu}=(-q^2g_{\mu\nu}+q_\mu q_\nu)\pi(q^2)$	photon proper self-energy
$D_F'(q)_{\mu\nu}$	full unrenormalized photon propagator
$\tilde{D}_F'(q)_{\mu\nu}$	full renormalized photon propagator
$\pi_c(q^2)=\lim_{\Lambda\to\infty}[\pi(q^2)-\pi(\mu^2)]$	subtracted photon proper self-energy
x	dimensionless variable $-q^2/m^2$
m_0, Z_2	electron bare mass and wave-function renormalization constant
$\beta(\alpha)$	coefficient of $\partial/\partial\alpha$ in the Callan-Symanzik equation
$f(\alpha_0)$	coefficient of the logarithmically divergent part of the photon proper self-energy in the JBW model
j_μ	electromagnetic current operator
$\tilde{S}_F'(p)$	renormalized electron propagator
$\gamma(\alpha), \delta(\alpha)$	coefficient functions appearing in the Callan-Symanzik equation for the electron propagator
$\pi_c^{[n]}, \beta^{[n]}$	parts of π_c, β with exactly n closed fermion loops
$T_{2n}^{[1]}=T_{\mu_1\cdots\mu_{2n}}^{[1]}(q_1,\ldots,q_{2n};m,y)$	single-fermion-loop $2n$-point function ($n\geq 2$) with coupling y
$\pi_{2n}^{[1]}$	modified 2-point function defined as a contraction on $T_{2n}^{[1]}$

$\pi_c^{[n,r]}$	part of $\pi_c^{[n]}$ in which exactly r closed fermion loops remain when all internal photon self-energy parts are shrunk down to points
$w(\rho/m^2, \alpha)$	Källén-Lehmann spectral function for the photon propagator
$F_B^{[1]}(y)$	coefficient of the logarithmically divergent part of the sum of single charged boson loop vacuum polarization diagrams
η	combination $\alpha(\xi-1)$ through which gauge dependence occurs

APPENDIX B: CALLAN-SYMANZIK EQUATIONS AND APPLICATION TO THE ELECTRON PROPAGATOR

In this Appendix we derive the Callan-Symanzik equations for massive photon (i.e., infrared cutoff) spinor electrodynamics in an arbitrary covariant gauge. We are particularly interested in the equations for the electron propagator and the electron-scalar vertex, which can be used to derive the JBW asymptotic form for the electron propagator discussed in Sec. IID2. To begin, we recall that the gauge parameter ξ enters into the theory only via the quantity $\alpha_b D_F^0(q)_{\mu\nu}$, which according to Eqs. (2) and (7b) can be written as

$$\alpha_b D_F^0(q)_{\mu\nu} = \alpha_b\left(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2}\right)\frac{1}{q^2-\mu_0^2}\,\frac{-\Lambda^2}{q^2-\Lambda^2} + \alpha(\xi-1)\frac{q_\mu q_\nu}{q^2}\,\frac{1}{q^2-\mu^2}\,\frac{-\Lambda^2}{q^2-\Lambda^2}\,. \tag{B1}$$

In particular, we see that ξ always appears in the combination $\alpha(\xi-1)$, a fact which we exploit by displaying the arguments of the renormalized electron propagator $\tilde S_F'^{-1}$ and the electron wave function renormalization Z_2 in the form

$$\begin{aligned}\tilde S_F'^{-1} &= \tilde S_F'^{-1}[p,\mu,m,\alpha,\eta]\,,\\ Z_2 &= Z_2[\Lambda,\mu,m,\alpha,\eta]\,,\\ \eta &= \alpha(\xi-1)\,.\end{aligned} \tag{B2}$$

To derive the Callan-Symanzik equations for the electron propagator, we start by writing down the equation connecting the renormalized and unrenormalized electron propagators,

$$\begin{aligned}\tilde S_F'[p,\mu,m,\alpha,\eta] &= Z_2 S_F'^{-1}\\ &= Z_2[\Lambda,\mu,m,\alpha,\eta](\not p - m_0 - \Sigma)\,,\end{aligned} \tag{B3}$$

with Σ the electron proper self-energy part. We now make independent variations in the physical electron and photon masses m and μ, keeping Λ and α_b fixed and simultaneously making a gauge transformation which keeps $\eta=\alpha(\xi-1)$ fixed. These variations are described by the respective differential operators

$$\begin{aligned} m\frac{d}{dm} &= m\frac{\partial}{\partial m}+\beta_m\alpha\frac{\partial}{\partial\alpha}\,,\\ \mu\frac{d}{d\mu} &= \mu\frac{\partial}{\partial\mu}+\beta_\mu\alpha\frac{\partial}{\partial\alpha}\,,\end{aligned} \tag{B4}$$

with β_m and β_μ defined by

$$\begin{aligned}\alpha\beta_m &= m\frac{d\alpha}{dm} = Z_3^{-1} m\frac{d}{dm}Z_3\,,\\ \alpha\beta_\mu &= \mu\frac{d\alpha}{d\mu} = Z_3^{-1}\mu\frac{d}{d\mu}Z_3\,,\end{aligned} \tag{B5}$$

in analogy with Eq. (42). Applying these differential operators to Eq. (B4), and observing that the unrenormalized propagator $\not p - m_0 - \Sigma$ depends on m and μ implicitly through its dependence on m_0 and μ_0 and explicitly through the factor $1/(q^2-\mu^2)$ in the gauge term, we get the Callan-Symanzik equations for the electron propagator,

$$\begin{aligned}\left(m\frac{\partial}{\partial m}+\alpha\beta_m\frac{\partial}{\partial\alpha}+\gamma_m\right)\tilde S_F'^{-1} &= -(1+\delta_m)\bar\Gamma_S+\mu^2\alpha\beta_m\bar\Gamma_{S'}\,,\\ \left(\mu\frac{\partial}{\partial\mu}+\alpha\beta_\mu\frac{\partial}{\partial\alpha}+\gamma_\mu\right)\tilde S_F'^{-1} &= -\delta_\mu\bar\Gamma_S+\mu^2(\alpha\beta_\mu-2)\bar\Gamma_{S'}+2\mu^2(\xi-1)\bar\Gamma_{S''}\,.\end{aligned} \tag{B6}$$

In writing this equation we have introduced the following additional definitions:

$$\begin{aligned} Z_2^{-1} m\frac{d}{dm}Z_2 &= -\gamma_m\,,\\ Z_2^{-1}\mu\frac{d}{d\mu}Z_2 &= -\gamma_\mu\,,\\ m_0^{-1} m\frac{d}{dm}m_0 &= 1+\delta_m\,,\\ m_0^{-1}\mu\frac{d}{d\mu}m_0 &= \delta_\mu\,,\\ \bar\Gamma_S &= m_0 Z_2\left(1+\frac{\partial\Sigma}{\partial m_0}\right),\end{aligned} \tag{B7}$$

$$\bar{\Gamma}_{S'} = \frac{Z_2}{Z_3}\frac{\partial\Sigma}{\partial\mu_0^2},$$

$$\bar{\Gamma}_{S''} = Z_2 \Gamma_{S''}.$$

The vertex part $\Gamma_{S''}$ is defined as the sum of terms in which each internal photon propagator $\alpha_b D_F^0(q)_{\mu\nu}$ is replaced in succession by

$$\alpha(q_\mu q_\nu/q^2)(q^2-\mu^2)^{-2}[-\Lambda^2/(q^2-\Lambda^2)]. \tag{B8}$$

Note that the derivative $\partial/\partial\alpha$ in Eq. (B6) acts only on the α dependence explicitly displayed in Eq. (B2) and not on the α dependence which is implicitly present as a result of the dependence on η. Let us now simplify Eq. (B6) in two ways. First we pass to the region of asymptotic $-p^2/m^2$, which allows us to drop the terms $\bar{\Gamma}_{S'}$ and $\bar{\Gamma}_{S''}$ on the right-hand side, since these vanish asymptotically. Secondly, we observe that we are really only interested in keeping the infrared cutoff μ^2 where it appears in divergent terms proportional to a power of $\ln\mu^2$. We get these divergent terms correctly even if we neglect those terms in Eq. (B6) which vanish as $O(\mu^2(\ln\mu^2)^n)$ as $\mu^2\to 0$. Making these simplifications and adding the second equation in Eq. (B6) to the first gives the desired form of the Callan-Symanzik equations for the asymptotic electron propagator,

$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)\right]\tilde{S}_F'^{-1}\sim-[1+\delta(\alpha)]\tilde{\Gamma}_S,$$
$$\left[\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right]\tilde{S}_F'^{-1}\sim 0, \tag{B9}$$

where[41]

$$\beta(\alpha)=\beta_m|_{\mu^2=0},$$
$$\delta(\alpha)=\delta_m|_{\mu^2=0}, \tag{B10}$$
$$\gamma(\alpha,\eta)=(\gamma_m+\gamma_\mu)|_{\mu^2=0}.$$

A precisely analogous procedure[18] yields the Callan-Symanzik equations for the asymptotic electron-scalar vertex,

$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)-\delta(\alpha)\right]\tilde{\Gamma}_S\sim 0,$$
$$\left[\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right]\tilde{\Gamma}_S\sim 0. \tag{B11}$$

Finally, in the limit as $\mu^2\to 0$ Eq. (B7) for Z_2 and m_0 can be rewritten in the form

$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)\right]Z_2=0,$$
$$\left[\mu\frac{\partial}{\partial\mu}+\gamma_\mu\right]Z_2=0, \tag{B12}$$
$$\left[m\frac{\partial}{\partial m}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}-\delta(\alpha)\right]\left(\frac{m_0}{m}\right)=0,$$

closely analogous to Eq. (45b) for Z_3 given in the text.

Let us now use Eqs. (B9)–(B12) to study the asymptotic behavior of $\tilde{S}_F'$ and the large $-\Lambda$ behavior of m_0 and Z_2 in the cases of type-1 and type-2 asymptotic behavior (cf. Sec. II B).

Type 1. In this case the physical coupling α is equal to the value α_1 which satisfies $q(\alpha_1)=\alpha_0$, $\beta(\alpha_1)=0$ and, as shown in Sec. II C, the asymptotic renormalized photon propagator αd_c^∞ is exactly equal to α_0. Because $\beta(\alpha)=0$, the $\partial/\partial\alpha$ terms disappear from Eqs. (B9)–(B12), and so these equations become the simplified Callan-Symanzik equations used in the analysis of Ref. 17 (apart from the change that the asymptotic coupling α_0 used in Ref. 17 is replaced now by the physical coupling $\alpha=\alpha_1$). For the asymptotic behavior of $\tilde{S}_F'(p)$ and the large $-\Lambda$ behavior of m_0 and Z_2 we thus get the scaling expressions of Eq. (66). Furthermore, we find the gauge transformation properties derived in Ref. 17 to be in accord with the conclusion which we have reached above, that the gauge parameter ξ appears only in the combination $\eta=\alpha(\xi-1)$.

Type 2. In this case $\alpha\neq\alpha_1$ and so $\beta(\alpha)\neq 0$. We proceed to analyze the asymptotic behavior under the conventional assumption that α_0 is a simple zero, and a point of regularity, of the Gell-Mann-Low function ψ, or equivalently[42] [cf. Eq. (53)] that α_1 is a simple zero and a point of regularity of β. As we have seen in Eqs. (32)–(35), this assumption corresponds to power law vanishing of the nonasymptotic piece h of the renormalized photon propagator. To study Eqs. (B9)–(B11) for $\tilde{S}_F'$ and $\tilde{\Gamma}_S$, we separate out the γ-matrix structure by writing

$$\tilde{S}_F'^{-1}=\not{p}F+mG,$$
$$m\tilde{\Gamma}_S=\not{p}H+mJ, \tag{B13}$$

which gives the equations

$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)\right]F\sim 0,$$
$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)\right]mG\sim-[1+\delta(\alpha)]mJ, \tag{B14}$$
$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)-\delta(\alpha)\right]J\sim 0.$$

The first of these three differential equations has the general integral

$$F=\exp\left[-\int_{L_F}^{\alpha}\frac{dz\,\gamma(z,\eta)}{z\beta(z)}\right]$$

$$\times\Phi_F\left[\ln\left(\frac{-p^2}{m^2}\right)+\int_{L_F}^{\alpha}\frac{2dz}{z\beta(z)},\ \mu^2/m^2,\eta\right],\qquad\text{(B15)}$$

with $\Phi_F[u,\mu^2/m^2,\eta]$ an arbitrary function of its arguments. Let us now consider the behavior of Eq. (B15) as $\alpha\to\alpha_1$. Since β has a zero at $z=\alpha_1$, the argument of the exponential prefactor and the argument u of the function Φ_F both become infinite. The only way for the function F to remain regular at $\alpha=\alpha_1$ is for the singularities of the exponential and of Φ_F at $\alpha=\alpha_1$ to precisely cancel. This can happen only if Φ_F has the following asymptotic behavior as u becomes infinite,

$$\Phi_F[u,\mu^2/m^2,\eta]\underset{u\to\infty}{\sim}C_F(\mu^2/m^2,\alpha_1,\eta)$$

$$\times\exp[\tfrac{1}{2}\gamma(\alpha_1,\eta)u]\,.\qquad\text{(B16)}$$

If we assume Eq. (B16), then when α is near α_1 we get

$$F\approx\exp\left[\int_{L_F}^{\alpha}dz\frac{\gamma(\alpha_1,\eta)-\gamma(z,\eta)}{z\beta(z)}\right]\times\text{finite terms},\qquad\text{(B17)}$$

which is regular because β vanishes with only a simple zero at $z=\alpha_1$. Let us now consider what happens as $-p^2/m^2$ becomes infinite, with α fixed at its physical value, different from α_1. Again u becomes infinite, this time because of the term $\ln(-p^2/m^2)$ in Eq. (B15), and so invoking Eq. (B16) gives us

$$F\sim C_F(\mu^2/m^2,\alpha_1,\eta)\exp\left[\int_{L_F}^{\alpha}\frac{\gamma(\alpha_1,\eta)-\gamma(z,\eta)}{z\beta(z)}\right]$$

$$\times\left(\frac{-p^2}{m^2}\right)^{\gamma(\alpha_1,\eta)/2}.\qquad\text{(B18)}$$

Thus, we see that even when $\alpha\neq\alpha_1$, *in the asymptotic limit F exhibits scaling behavior with a scaling exponent γ characteristic of the value α_1 at which β vanishes.*[43] An identical argument can be used to integrate the equations for G and J in Eq. (B14) and those for Z_2 and m_0/m in Eq. (B12), and finally the equation

$$\mu\frac{\partial}{\partial\mu}(\tilde{S}_F'^{-1}/Z_2)=\frac{1}{Z_2^{\,2}}\left(Z_2\mu\frac{\partial}{\partial\mu}\tilde{S}_F'^{-1}-\tilde{S}_F'^{-1}\mu\frac{\partial}{\partial\mu}Z_2\right)$$

$$\sim 0\qquad\text{(B19)}$$

can be used to relate the μ dependence of the resulting constants of integration. The procedure unfolds in complete analogy[44] with the treatment of the JBW model given in Ref. 17, and the results obtained are of the same form as in Eq. (66), apart from the more complex structure of the integration constants seen in Eq. (B18).

To conclude, we reemphasize that in order to derive Eq. (B18), we need the twin assumptions of a simple zero in β and of regularity of the theory around α_1. If β vanishes more rapidly than with a simple zero at α_1, the exponential factor in Eq. (B17) is still not regular at α_1, and so the argument for requiring Φ_F to have the particular asymptotic form given in Eq. (B16) is no longer compelling. For an alternative derivation of scaling behavior of the asymptotic electron propagator, which may be valid even when ψ (or equivalently, β) has a higher-order zero, see Sec. III C of the text.

---

[1]Particle Data Group, Rev. Mod. Phys. 43, S1 (1971).

[2]There are already stringent limits on the possible variation of α on a cosmological time scale. See P. J. Peebles and R. H. Dicke, Phys. Rev. 128, 2006 (1962); F. J. Dyson, Phys. Rev. Letters 19, 1291 (1967); A. Peres, *ibid.* 19, 1293 (1967); J. N. Bahcall and M. Schmidt, *ibid.* 19, 1294 (1967).

[3]See, for example, S. Weinberg, Phys. Rev. Letters 19, 1264 (1967); 27, 1688 (1971); L. D. Landau, in *Niels Bohr and the Development of Physics* (McGraw-Hill, New York, 1955), p. 60; A. Salam, paper presented at the Fifteenth International Conference on High-Energy Physics, Kiev, U.S.S.R., 1970 (unpublished).

[4]See the historical footnote on p. 607 of S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, Evanston, Ill., 1961).

[5]R. Jost and J. M. Luttinger, Helv. Phys. Acta 23, 201 (1950).

[6]M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954).

[7]K. Johnson, M. Baker, and R. Willey, Phys. Rev. 136, B111 (1964); K. Johnson, R. Willey, and M. Baker, *ibid.* 163, 1699 (1967); M. Baker and K. Johnson, *ibid.* 183, 1292 (1969); M. Baker and K. Johnson, Phys. Rev. D 3, 2516 (1971); 3, 2541 (1971).

[8]P. G. Federbush and K. Johnson, Phys. Rev. 120, 1296 (1960). See also R. Jost, in *Lectures on Field Theory and the Many-Body Problem*, edited by E. R. Caianiello (Academic, New York, 1961); K. Pohlmeyer, Commun. Math. Phys. 12, 204 (1969). The usual derivation of the Federbush-Johnson theorem depends heavily on assumptions of positivity and locality. As a result, the derivation cannot be directly applied to quantum electrodynamics, where, if one quantizes in the Lorentz gauge to guarantee locality, one must use the Gupta-Bleuler negative metric, while if one quantizes in the radiation gauge to

guarantee positivity, one loses local commutativity. It appears that this problem can be circumvented, and that a satisfactory proof for the case of electrodynamics can be given [F. Strocchi (unpublished)], at least in the charge-zero sector.

[9]B. Simon, in Proceedings of the 1972 Coral Gables Conference (unpublished), conjectures that field theories, as a function of coupling constant, will prove to be Borel summable. This would make it possible to determine them uniquely from a knowledge of their formal power series expansions. A cautionary remark is necessary here: As discussed by Simon, Borel summability of a function such as $F^{[1]}(y)$ cannot be deduced from the formal power-series coefficients alone – additional information about $F^{[1]}(y)$ is needed. In the same article, Simon also discusses a possible connection of Borel summability with the conjecture of F. J. Dyson, Phys. Rev. 85, 631 (1952), concerning singularities in field theory at zero coupling constant. However, one should note that in the specific examples of nonanalytic behavior given by Simon, the nonanalyticity arises from mass renormalization, not from the fermion vacuum polarization effects discussed both by Dyson and by us in the present paper.

[10]C. G. Callan, Phys. Rev. D 2, 1541 (1970); K. Symanzik, Commun. Math. Phys. 18, 227 (1970); and in *Springer Tracts in Modern Physics*, edited by G. Höhler (Springer, Berlin, 1971), Vol. 57, p. 222.

[11]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965). Throughout this paper, we follow wherever possible the Bjorken-Drell notation and metric conventions.

[12]Note that this is *not* just a "leading logarithm" approximation: *All* terms in each order of perturbation theory which do not vanish asymptotically are retained.

[13]A discussion of effects of dropping this assumption has been given by M. Astaud and B. Jouvet, Nuovo Cimento 63A, 5 (1969), and M. Astaud, *ibid.* 66A, 111 (1970).

[14]However, when $\alpha_w d_c[x, w, \alpha_w]$ is reexpanded as a power series in α, there will be an m dependence, introduced by the appearance of the nonasymptotic point zero in Eq. (18).

[15]T. Kinoshita, J. Math. Phys. 3, 650 (1962); T. D. Lee and M. Nauenberg, Phys. Rev. 133, B1549 (1964). See also A. Sirlin, Phys. Rev. D 5, 436 (1972), who discusses the connection between the infrared cancellation theorems and the Callan-Symanzik equations.

[16]M. Baker and K. Johnson, Phys. Rev. 183, 1292 (1969).

[17]We follow the approach of S. Coleman (unpublished) and A. Sirlin, Ref. 15.

[18]S. L. Adler and W. A. Bardeen, Phys. Rev. D 4, 3045 (1971). The quantity called $\delta(\alpha)$ in Eq. (41) is denoted by α in the work of Adler and Bardeen.

[19]S. Weinberg, Phys. Rev. 118, 838 (1960).

[20]See also K. Pohlmeyer, Commun. Math. Phys. 12, 204 (1969). Note that although the $2n$-point current correlation functions vanish for $n \geq 2$, the Federbush-Johnson theorem does *not* require the vanishing of the dispersive part of the photon proper self-energy (the case $n=1$), which would imply that $\alpha_0 = \alpha$. A heuristic way of understanding this is to note that the annihilation of the vacuum by the electromagnetic current implies the vanishing of the absorptive part of the general $2n$-point current correlation function ($n \geq 1$). This implies that in momentum space the correlation function must be a polynomial function of its four-momentum arguments. Gauge invariance tells us that this polynomial must contain as factors the four-momenta $k_1 \cdots k_{2n}$ of the $2n$ photons, and hence contains only terms of degree $2n$ or higher. On the other hand, Weinberg's theorem tells us that the amplitude behaves at worst as $(\text{momentum})^{4-2n} \times$ logarithms as all four-momenta are scaled to infinity. Thus the minimum degree $2n$ of the polynomial must satisfy the inequality $2n \leq 4 - 2n$, which is compatible only with $n=1$. Hence the $2n$-point current correlation function with $n \geq 2$ must vanish, while the two-point function is a polynomial of the form $(-q^2 g_{\mu\nu} + q_\mu q_\nu) \times$ constant, consistent with our initial assumption that the photon propagator αd_c is equal to the asymptotic value $\alpha_0 \neq \alpha$.

[21]It is still possible, of course, that the converse result is true, and might be provable if the analyticity structure of the single-fermion-loop diagram as a function of the external photon four-momenta is taken into account.

[22]See also S. L. Adler and W. A. Bardeen, erratum to Ref. 18 (to be published).

[23]Equation (77) is the generalization of an identity, due to M. L. Goldberger, which formally relates an integral over the virtual Compton amplitude to the coupling constant derivative of the electron self-energy part in quantum electrodynamics. I am grateful to R. F. Dashen for reminding me of that identity and for suggesting its applicability to vacuum polarization loops.

[24]This formula also gives the correct fractional weights for vacuum diagrams (the case $n=0$, $j \geq 1$).

[25]Even though Eq. (88) is true only by virtue of taking a nonperturbative sum of diagrams to infinite order, we are *assuming*, with no attempt at justification, that the difference between Eq. (89) and Eq. (88) vanishes for small m as it would in perturbation theory. The proportionality to m^2, rather than just to m, follows from the fact that because of charge conjugation invariance $T^{[1]}$ is an even function of m.

[26]In writing Eq. (90) we make no commitment as to the size of the mass cm^2 which effectively cuts off the integral – it could in principle be exceedingly large, since it arises from a cancellation of asymptotically dominant terms which involves all orders of perturbation theory. Experimental tests of electrodynamics which are sensitive to vacuum polarization effects could be used to set a lower limit on the possible value of cm^2.

[27]One might ask why, even for $y \neq \alpha_0$, one cannot simply add and subtract

$$T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; 0, y)$$

in the integrand of Eq. (76), thus leading to the following modified version of Eq. (85),

$$[\pi^{(1)}_{2n}(q^2; m, y) - \pi^{(1)}_{2n}(q^2; m', y)](-q^2 g_{\mu\nu} + q_\mu q_\nu) = \tilde{I}_m - \tilde{I}_{m'} ,$$

with

$$\tilde{I}_m = \int \frac{d^4 q_1}{(2\pi)^4} \cdots \frac{d^4 q_{n-1}}{(2\pi)^4} \left(-\frac{i g^{\mu_1\mu_2}}{q_1^2}\right) \cdots \left(-\frac{i g^{\mu_{2n-3}\mu_{2n-2}}}{q_{n-1}^2}\right)$$
$$\times [T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; m, y)$$
$$- T^{[1]}_{\mu_1\mu_2\cdots\mu_{2n-3}\mu_{2n-2}\mu\nu}(q_1, -q_1, \ldots, q_{n-1}, -q_{n-1}, q, -q; 0, y)] .$$

The answer is that although $\tilde{I}_m$ is now ultraviolet convergent, the subtraction term makes $\tilde{I}_m$ a logarithmically divergent integral in the infrared of the general type

$$\int_0^\infty d\rho\left(\frac{1}{\rho+m^2}-\frac{1}{\rho}\right)=\int_0^\infty \frac{-m^2}{\rho(\rho+m^2)}.$$

Thus, the modified version of Eq. (85) is still an ambiguous expression of the form $\infty-\infty$. The significance of the special condition, Eq. (74), which holds when $y=\alpha_0$ is that it improves the ultraviolet behavior of I_m without simultaneously making the infrared behavior worse. This feature has been incorporated in the illustrative example given in Eq. (90).

[28]After this work was completed, we learned that K. Johnson (unpublished) knew a related argument suggesting a zero of infinite order in $\psi(y)$, obtained by working directly with the modified skeleton expansion described in Sec. II D1.

[29]Equations (99) and (102) clearly illustrate the distinction between type-1 and type-2 asymptotic behavior. When $\alpha_0\neq q(\alpha)$, the asymptotic behavior is type 2, and Eqs. (99) and (102) can both be developed as power series in $\ln x$. When $\alpha_0=q(\alpha)$, Eqs. (99) and (102) both degenerate to $\alpha d_c^\infty=\alpha_0$, as expected for type-1 asymptotic behavior.

[30]The discussion which follows depends only on the fact that the Gell-Mann–Low function vanishes with a zero of infinite order, and does not hinge crucially on the choice of Eq. (103) for h. To see this, we note from Table I that if the Gell-Mann–Low function vanishes with a zero of *finite* order N, the function $h(x,\alpha)$ vanishes asymptotically as $(\ln x)^{-1/(N-1)}$. Consequently, h^n vanishes as $(\ln x)^{-n/(N-1)}$ and the integral in Eq. (106) diverges for all $n\leq N-1$. Letting $N\to\infty$, we learn that in the case of an infinite-order zero of the Gell-Mann–Low function, the integral in Eq. (106) diverges for all n. Note that in making the distinction between the case where h vanishes as a power of x and the case where h vanishes more slowly, it is important to adhere to our convention of "vacuum-polarization-insertion-wise" summation, which requires us to sum the logarithmic series defining d_c^∞ *before* passing to the asymptotic limit. This is particularly important in the case of Eq. (102), where the logarithmic series has only a finite radius of convergence and so cannot be used to describe the asymptotic region.

[31]W. A. Bardeen (unpublished).

[32]The question of whether the Federbush-Johnson theorem can be extended outside the charge-zero sector in electrodynamics is an important one and deserves further study. If it can be extended sufficiently to imply the vanishing of the electron-photon vertex part, then using the Ward identity to relate the vertex part to the asymptotic electron propagator in Eq. (66) implies that

$$\lim_{m\to 0} F_1 C_1 m^{-\gamma}=0$$

when the eigenvalue condition is satisfied. If $F_1\neq 0$, this equation then tells us that Z_2 must vanish.

[33]G. Källén, Helv. Phys. Acta 25, 417 (1952); H. Lehmann, Nuovo Cimento 11, 342 (1954).

[34]Note that there is no contradiction between the fact that $\alpha_6>\alpha$ and the assertion, essential to the argument of Sec. II D1, that the spectral function vanishes as $m^2\to 0$. For illustrative purposes let us follow the example of Eq. (90) and take

$$w(x,\alpha)=c/[(1+x)(1+cx)]$$

with $c>0$. Then Eq. (129) becomes

$$\alpha_0=\alpha+\int_0^\infty\frac{c\alpha m^2 d\rho}{(\rho+m^2)(\rho+cm^2)}=\alpha+c\alpha\frac{\ln c}{c-1}>\alpha,$$

but *for fixed* ρ the spectral function (i.e., the integrand) vanishes as $m^2\to 0$.

[35]The fact that y_0 appears as an *infinite-order* zero of $F^{[1]}$ means that it may be possible to determine y_0 from the limiting behavior of the nth term in the perturbation expansion for $F^{[1]}$ as $n\to\infty$. For example, suppose that $F^{[1]}$ actually has a convergent power series expansion around $y=0$ with radius of convergence y_0,

$$F^{[1]}(y)=\sum_{n=0}^\infty c_n y^n.$$

Then y_0 is given by the limit formula

$$y_0=\lim_{n\to\infty}|c_n|^{1/n}.$$

Since c_n describes the fermion loop with n internal virtual photons, it is thus conceivable that y_0 can be computed in a semiclassical (large-photon-number) calculation.

[36]Dyson (Ref. 9) also considers the alternative possibility, that electrodynamics by itself is not a complete theory, and becomes consistent only when other interactions are taken into account.

[37]The sixth-order result for $F^{[1]}$ is due to J. L. Rosner, Phys. Rev. Letters 17, 1190 (1966), and Ann. Phys. (N.Y.) 44, 11 (1967). The fourth-order expansion for $F_B^{[1]}(y)$ is due to Z. Białynicka-Birula, Bull. Acad. Polon. Sci. 13, 369 (1965). Conflicting results in the fourth-order boson calculation have been claimed by I.-J. Kim and C. R. Hagen, Phys. Rev. D 2, 1511 (1970). However, D. Sinclair (unpublished) has located an error in the work of Kim and Hagen which, when corrected, gives Białynicka-Birula's result. Sinclair has also rechecked this result independently by Rosner's method of calculation.

[38]For a review of this point of view, see D. J. Gross and S. B. Treiman, Phys. Rev. D 4, 1059 (1971).

[39]For an alternative explanation, which regards scaling as an intermediate energy manifestation of compositeness of the nucleon, see S. D. Drell and T. D. Lee, Phys. Rev. D 5, 1738 (1972).

[40]Since $F_B^{[1]}(y)$ is different from $F^{[1]}(y)$ our scheme could, in principle, accommodate a fractionally charged elementary boson.

[41]The renormalization constants m_0 and Z_3 are gauge-invariant and are also infrared finite as $\mu^2\to 0$. These properties account, respectively, for the facts that $\delta(\alpha)$ and $\beta(\alpha)$ are independent of η and that δ_μ and β_μ vanish as $\mu^2\to 0$. In Ref. 18, it is shown that the gauge dependence of $\gamma(\alpha,\eta)$ is strictly additive, i.e., $\gamma(\alpha,\eta)-\gamma(\alpha,\eta')=(\eta-\eta')/(2\pi)$.

[42]We assume that the mapping $q(\alpha)$ is well behaved near α_1, in particular, that $q'(\alpha_1)\neq 0$.

[43]This was first pointed out by K. G. Wilson, Phys. Rev. D 3, 1818 (1971). Our treatment is suggested by the procedure of C. G. Callan, Ref. 10.

[44]In particular, comparison of the second and third equations in Eq. (B14) shows that $G=-J$ is a particular integral of the differential equation for G, and a simple application of Weinberg's theorem shows that one cannot add a solution of the homogeneous equation

$$\left[m\frac{\partial}{\partial m}+\mu\frac{\partial}{\partial\mu}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}+\gamma(\alpha,\eta)\right]mG=0$$

to the particular solution.

PHYSICAL REVIEW D VOLUME 6, NUMBER 10 15 NOVEMBER 1972

Constraints on Anomalies*

Stephen L. Adler and Curtis G. Callan, Jr.†
Institute for Advanced Study, Princeton, New Jersey 08540

and

David J. Gross
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540

and

Roman Jackiw
Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 16 June 1972)

The various coupling-constant-dependent numbers describing anomalous commutators are constrained by the nonrenormalization of the axial-vector–current anomaly. The axial-vector current continues to behave anomalously even if the underlying unrenormalized field theory is finite due to the vanishing of the Gell-Mann–Low eigenvalue function.

I. INTRODUCTION

It has now been established that the canonical formalism of quantum field theory frequently yields results that are not verified in perturbation theory.[1] These "anomalies" are of two distinct kinds. Firstly there are failures of the Bjorken-Johnson-Low (BJL)[2] limit: Equal-time commutators between operators, when evaluated by the BJL technique in perturbation theory, usually do not agree with the canonical determination of these commutators. A well-known consequence is the failure of the Callan-Gross sum rule for electroproduction.[3] Secondly there are violations of Ward identities associated with exact or partial symmetries; the two known examples being the triangle anomaly of the axial-vector current and the trace anomaly of the new improved energy-momentum tensor.[1] (When a Ward identity is anomalous, there is also a corresponding BJL anomaly.) The Sutherland-Veltman low-energy theorem for neutral pion decay is falsified as a consequence.[4] Both categories of anomalies arise from the divergences of unrenormalized perturbation theory, which require the introduction of regulators to define the theory. The BJL anomaly reflects the noncommutativity of the BJL high-energy limit with the infinite regulator limit which must be taken to define renormalized, physical amplitudes. Failures of Ward identities arise when no regulator

exists which preserves the relevant symmetry.

Although the common cause for both classes of anomalies is evident, it has not been appreciated that an intimate relationship exists between the BJL anomalies and the failures of Ward identities. In this paper we demonstrate that the very interesting analysis by Crewther[5] of the triangle anomaly in terms of Wilson's short-distance expansion[6] can be extended to exhibit this relationship. Further we show that in lowest nontrivial order of perturbation theory the q-number anomaly in the equal-time commutator of space-components of currents can be completely determined in terms of the c-number anomaly in the equal-time commutator between the time component and space component of the current, i.e., the ordinary Schwinger term. Finally we inquire to what extent the canonical formalism can be reestablished if the unrenormalized theory becomes finite due to the vanishing of the Gell-Mann–Low[7] eigenvalue function. Our conclusion, at least for the axial-vector current, is that naive manipulations continue to lead to error.

II. THE CREWTHER ANALYSIS

Assume that one is dealing with a theory which is conformally invariant at short distances. Consider the vector–vector–axial-vector current amplitude

$$T^{\mu\nu\alpha}_{abc}(x,y,z)=\langle 0|T(V^\mu_a(x)V^\nu_b(y)A^\alpha_c(z))|0\rangle . \tag{2.1}$$

Schreier[8] has shown that a conformally invariant three-index pseudotensor of dimension 9 must be proportional to the fermion triangle graph constructed from massless fermions in free field theory. Hence (2.1) is given by[9]

$$T^{\mu\nu\alpha}_{abc}(x,y,z)=N\Delta^{\mu\nu\alpha}_{abc}(x,y,z)+\cdots , \tag{2.2a}$$

$$\Delta^{\mu\nu\alpha}_{abc}(x,y,z)=\frac{d_{abc}}{16\pi^6}\,\frac{\mathrm{Tr}\,\gamma^5\gamma^\mu\gamma^\delta\gamma^\nu\gamma^\epsilon\gamma^\alpha\gamma^\varphi(x-y)_\delta(y-z)_\epsilon(z-x)_\varphi}{[(x-y)^2-i\epsilon]^2[(y-z)^2-i\epsilon]^2[(z-x)^2-i\epsilon]^2}\,. \tag{2.2b}$$

Here N is a number and the dots in (2.2a) represent less singular, non-scale-invariant contributions to $T^{\mu\nu\alpha}_{abc}(x,y,z)$, which vanish in the scale-invariant (= conformally invariant) limit. The precise assumption about these subdominant terms is that they can be identified and separated from $\Delta^{\mu\nu\alpha}_{abc}(x,y,z)$ in *sequential* short-distance limits,

$$\lim_{x_i\to x_k}\ \lim_{x_j\to x_k} T^{\mu\nu\alpha}_{abc}(x,y,z)=\lim_{x_i\to x_k}\ \lim_{x_j\to x_k} N\Delta^{\mu\nu\alpha}_{abc}(x,y,z)+\text{less singular terms}, \tag{2.3}$$

where $\{x_i, x_j, x_k\}$ are any of $\{x, y, z\}$. Thus we know that

$$\lim_{x\to 0}\lim_{y\to 0} T^{\mu\nu\alpha}_{abc}(x,y,0)=\frac{Nd_{abc}}{4\pi^6}\,\frac{\epsilon^{\nu\alpha\delta\epsilon}y_\epsilon(2x^\mu x_\delta-g^\mu_\delta x^2)}{(y^2-i\epsilon)^2(x^2-i\epsilon)^4}+\text{less singular terms}, \tag{2.4a}$$

$$\lim_{z\to 0}\lim_{x\to 0} T^{\mu\nu\alpha}_{abc}(x,0,z)=\frac{Nd_{abc}}{4\pi^6}\,\frac{\epsilon^{\mu\nu\delta\epsilon}x_\epsilon(2z^\alpha z_\delta-g^\alpha_\delta z^2)}{(x^2-i\epsilon)^2(z^2-i\epsilon)^4}+\text{less singular terms}. \tag{2.4b}$$

Next a scale-invariant short-distance expansion for current commutators is postulated,

$$[V^\mu_a(x), V^\nu_b(0)] \underset{x\sim 0}{=} -iS_{VV}\delta_{ab}(g^{\mu\nu}x^2-2x^\mu x^\nu)\,\frac{\epsilon(x^0)\delta'''(x^2)}{6\pi^3}+iK_{VV}d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}\,A^\alpha_c(0)x^\beta\,\frac{\epsilon(x^0)\delta'(x^2)}{\pi}+\cdots , \tag{2.5a}$$

$$[A^\mu_a(x), A^\nu_b(0)] \underset{x\sim 0}{=} -iS_{AA}\delta_{ab}(g^{\mu\nu}x^2-2x^\mu x^\nu)\,\frac{\epsilon(x^0)\delta'''(x^2)}{6\pi^3}+iK_{AA}d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}\,A^\alpha_c(0)x^\beta\,\frac{\epsilon(x^0)\delta'(x^2)}{\pi}+\cdots , \tag{2.5b}$$

$$[V^\mu_a(x), A^\nu_b(0)] \underset{x\sim 0}{=} iK_{VA}d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}V^\alpha_c(0)x^\beta\,\frac{\epsilon(x^0)\delta'(x^2)}{\pi}+\cdots . \tag{2.5c}$$

The dots indicate less singular contributions, or operators with quantum numbers and symmetries different from the exhibited terms. S and K are constants which appear in the following equal-time commutators,

$$[V^0_a(x), V^i_b(0)]|_{x^0=0}=i\delta_{ab}S_{VV}\Lambda\partial^i\delta^3(\vec{x})-\frac{i}{24\pi^2}\delta_{ab}S_{VV}\partial^i\partial_k\partial^k\delta^3(\vec{x})+\cdots , \tag{2.6a}$$

$$[V^i_a(x), V^j_b(0)]|_{x^0=0}=iK_{VV}d_{abc}\epsilon^{ijk}A^k_c(0)\delta^3(\vec{x})+\cdots , \tag{2.6b}$$

etc.

Λ is a quadratically divergent constant, and the omitted terms have different quantum numbers. An expan-

sion similar to (2.5) is written for T products,

$$T(V_a^\mu(x)V_b^\nu(0)) \underset{x\sim 0}{=} S_{VV}\frac{\delta_{ab}(g^{\mu\nu}x^2-2x^\mu x^\nu)}{2\pi^4(x^2-i\epsilon)^4}-K_{VV}\frac{d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}A_c^\alpha(0)x^\beta}{2\pi^2[x^2-i\epsilon]^2}+\cdots, \tag{2.7a}$$

$$T(A_a^\mu(x)A_b^\nu(0)) \underset{x\sim 0}{=} S_{AA}\frac{\delta_{ab}(g^{\mu\nu}x^2-2x^\mu x^\nu)}{2\pi^4(x^2-i\epsilon)^4}-K_{AA}\frac{d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}A_c^\alpha(0)x^\beta}{2\pi^2[x^2-i\epsilon]^2}+\cdots, \tag{2.7b}$$

$$T(V_a^\mu(x)A_b^\nu(0)) \underset{x\sim 0}{=} -K_{VA}\frac{d_{abc}\epsilon^{\mu\nu}{}_{\alpha\beta}V_c^\alpha(0)x^\beta}{2\pi^2(x^2-i\epsilon)^2}+\cdots. \tag{2.7c}$$

Crewther's observation is that the constants N, S, and K are not independent.[5] From (2.1) and (2.7c) it follows that

$$\lim_{y\to 0} T_{abc}^{\mu\nu\alpha}(x,y,0) = -K_{VA}\frac{d_{bcd}\epsilon^{\nu\alpha}{}_{\omega\varphi}y^\varphi}{2\pi^2(y^2-i\epsilon)^2}\langle 0|T(V_a^\mu(x)V_d^\omega(0))|0\rangle \tag{2.8a}$$

while (2.6a) implies that

$$\lim_{x\to 0}\lim_{y\to 0} T_{abc}^{\mu\nu\alpha}(x,y,0) = S_{VV}K_{VA}\frac{d_{abc}\epsilon^{\nu\alpha\delta\epsilon}y_\epsilon(2x^\mu x_\delta-g_\delta^\mu x^2)}{4\pi^6(y^2-i\epsilon)^2(x^2-i\epsilon)^4}. \tag{2.8b}$$

Hence upon comparing (2.8b) and (2.4a), one finds

$$N=S_{VV}K_{VA}. \tag{2.9a}$$

Additionally, from (2.1), (2.4b), (2.7a), and (2.7b) it follows that

$$N=S_{AA}K_{VV}. \tag{2.9b}$$

If a similar analysis is performed on the axial-vector–axial-vector–axial-vector current amplitude one gets

$$N'=S_{AA}K_{AA}, \tag{2.10}$$

where N' is the proportionality constant defined analogously to (2.2a).[10] As emphasized by Crewther, the interest in relations (2.9) and (2.10) derives from the fact that S and K are measurable (in principle) in various deep-inelastic processes – thus the anomaly in low-energy processes, at an unphysical point, is directly determined by experimental high-energy behavior.

III. CONSTRAINTS ON ANOMALIES

We shall use (2.9) and (2.10) to probe the structure of anomalies in various models. Consider first a free massless field theory with

$$V_a^\mu(x)=:\bar\psi(x)\gamma^\mu\tfrac12\lambda_a\psi(x):,$$

$$A_a^\mu(x)=:\bar\psi(x)i\gamma^\mu\gamma_5\tfrac12\lambda_a\psi(x):.$$

It is trivial to verify that, as already stated by Crewther,[5] (2.9) and (2.10) are satisfied with $N=K=S=1$. The nonvanishing of S and N is conventionally described as anomalous. A naive evaluation of the equal-time commutator (2.6a) yields a vanishing result; the nonvanishing of S measures the famous Schwinger-term anomaly. Similarly a naive evaluation of $\partial_\mu^x T_{abc}^{\mu\nu\alpha}(x,y,z)$, $\partial_\nu^y T_{abc}^{\mu\nu\alpha}(x,y,z)$, and $\partial_\alpha^z T_{abc}^{\mu\nu\alpha}(x,y,z)$ yields zero since the currents are conserved. Nevertheless, as Schreier has shown,[8] one cannot consistently set all divergences of $\Delta_{abc}^{\mu\nu\alpha}(x,y,z)$ to zero, because this quantity is singular when all three points coincide. In momentum space this corresponds to the well-known violation of Ward-Takahashi identities of the fermion, axial-vector triangle graph. Hence N measures the axial-vector–current anomaly.[1] Since a naive determination of K from the equal-time commutator (2.6b) also gives $K=1$, we have a connection between the anomalies: $N=S$; the triangle anomaly is a consequence of the Schwinger-term anomaly.

Next consider a fermion theory with an SU(3)-invariant Yukawa interaction of strength g involving neutral vector gluons. (Spin-zero gluons render the axial-vector current infinite; hence we do not consider them.) In order to apply Crewther's analysis,[5] it is necessary to satisfy his hypotheses: (1) the existence of finite currents; (2) the existence of a scale-invariant expansion for products of currents, (2.5) and (2.7); and (3) the existence of a conformally invariant short-distance limit for $T_{abc}^{\mu\nu\alpha}(x,y,z)$, (2.2). No complete calculation of $T_{abc}^{\mu\nu\alpha}(x,y,z)$ in higher order has been performed which can be used to check the third hypothesis. We shall nonetheless *assume* that this result is valid, provided the other two are satisfied. This assumption is motivated by the fact that the triangle anomaly has no higher-order corrections,[11] and is almost certainly true for the class of graphs, discussed in detail below, which contain only a single fermion loop. (See the Appendix.) Therefore we set $N=N'=1$, *even in the presence of interactions*. In order to satisfy the first hypothesis, we must not consider SU(3) singlet vec-

tor currents, since these are not well defined in perturbation theory. The interaction with the vector gluon gives rise to infinite vacuum polarization, which modifies the singlet current.

In lowest-order perturbation theory in this model, a scale-invariant expansion for current products exists. This can be seen as follows. A BJL-limit determination of the equal-time commutator (2.6b) yields a finite expression.[1] Hence the q-number portion of the expansion (2.5) and (2.7) exists. That the c-number part also exists follows from the Jost-Luttinger calculation of the proper vacuum polarization tensor.[12] Their result is that in second order of perturbation theory this object is no more singular than in the free-field model. [In momentum space both the free-field graph and the lowest order graphs of Fig. 1 go as a single power of $\ln(-k^2)$ for large k.] However, S and K depart from their free-field values. The Jost-Luttinger formula[12] for S is $1+3g^2/16\pi^2$. Because $N=1$, we must have $K=(1+3g^2/16\pi^2)^{-1}\approx 1-3g^2/16\pi^2$; and the BJL calculation of K gives indeed this answer.[1] Hence we see that the BJL anomaly in the commutator of two spatial components of currents is determined by the higher-order terms in the c-number Schwinger term.

Beyond lowest order, perturbation theory no longer satisfies the hypotheses (1) and (2). The axial-vector current ceases to be well defined, since the graph of Fig. 2 is not rendered finite by external wave-function renormalization factors.[1] Also the c-number portion of the expansion (2.5) and (2.7) is not of the assumed scale-invariant form, since the proper vacuum polarization tensor acquires quadratic and higher powers of $\ln(-k^2)$. (No calculations have been performed on the q-number part of the expansion; but we expect that it too is no longer scale-invariant.) However, subsets of graphs can be chosen which probably continue to satisfy the hypotheses. For example if fermion creation and annihilation is ignored, then the vacuum expectation value of the currents is given by the one-fermion-loop graphs. In this approximation we have[13]

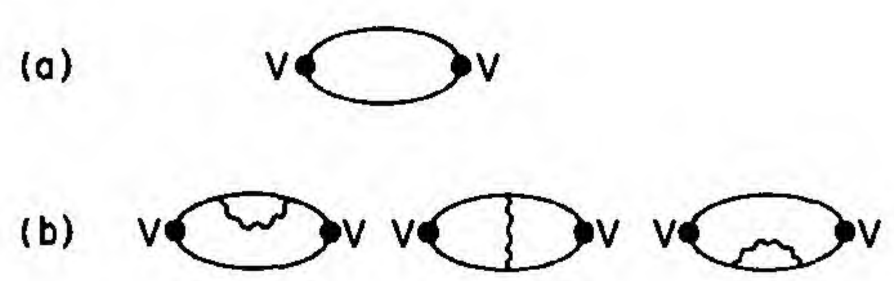


FIG. 1. Contributions to $\int d^4x\, e^{ikx}\langle 0|T(V_a^\mu(x)V_b^\nu(0))|0\rangle$ which go as $\ln(-k^2)$ for large k. (a) Free-field theory graph. (b) Lowest-order perturbation theory graphs.

$$\int\frac{d^4x}{(2\pi)^4}e^{iqx}\langle 0|T(V_a^\mu(x)V_b^\nu(0))|0\rangle^{\text{one-fermion-loop}}$$
$$=(g^{\mu\nu}q^2-q^\mu q^\nu)\frac{i}{24\pi^2}\delta_{ab}F(g^2)\ln(-q^2/m^2)$$
$$+\text{less singular terms}. \qquad (3.1\text{a})$$

Here $F(g^2)$ is the Baker-Johnson function,[13] whose first three terms in a power series expansion are known:

$$F(g^2)=1+\frac{3g^2}{16\pi^2}-\frac{3}{512}\frac{g^4}{\pi^4}+\cdots. \qquad (3.1\text{b})$$

Also the axial-vector current is no longer infinite, since the graph of Fig. 2 is absent. Evidently the c-number term in the expansion (2.5) and (2.7) is scale invariant with $S=F(g^2)$, and it is likely that so also is the q-number term. Hence we conjecture that if the current commutator were computed in the BJL limit, without including fermion creation or annihilation processes, one would find

$$K(g^2)=\frac{1}{F(g^2)}=1-\frac{3g^2}{16\pi^2}+\frac{21}{512}\frac{g^4}{\pi^4}+\cdots. \qquad (3.2)$$

IV. ANOMALIES IN THE GELL-MANN-LOW LIMIT

We consider quantum electrodynamics, and assume that the Gell-Mann-Low eigenvalue function possesses a zero, so that Z_3 is finite.[7] Now one *can* discuss currents, since the vacuum polarization no longer diverges. In this limit it should be possible to set the electron mass m to zero and scale invariance becomes exact.[14] (Z_2, the electron wave-function renormalization constant, can be made finite by appropriate choice of gauge.) We examine this (hypothetical) theory in the context of the ideas developed in Secs. II and III. It will be seen that singular behavior survives even in this finite theory and that naive canonical reasoning continues to be inapplicable.[15] [All our previous formulas hold with SU(3) indices suppressed, and the following replacements: $V_a^\mu - J^\mu$, $A_a^\mu - J_5^\mu$, $d_{abc} - 2$, $\delta_{ab} - 2$.] Observe first that, since the triangle anomaly has no radiative corrections, N continues to be equal to unity. However, because Z_3 is *finite*, the quadratically divergent Schwinger term is ab-

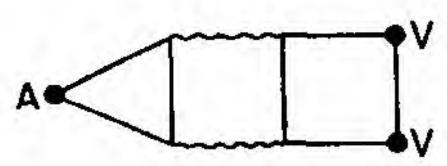


FIG. 2. Graph which renders the axial-vector current infinite.

sent; i.e., $S=0$. Since $K=1/S$, we see that the coefficient of the axial-vector current in the short-distance expansion of the product of two currents is infinite. In other words the c-number singularity (2.5) is weaker than x^{-6}, but the q-number singularity is stronger than x^{-3}, so that their product remains singular as x^{-9}. Consequently the BJL limit, which naively gives (2.5), is anomalous, and this is true regardless whether or not the electron mass is set to zero.

Further difficulties emerge if we set the electron mass to zero. In that limit all vacuum matrix elements of current products vanish by the Federbush-Johnson theorem,[14]

$$\langle 0|J^{\mu_1}(x_1)\cdots J^{\mu_n}(x_n)|0\rangle=0. \tag{4.1}$$

Nevertheless we now show that one cannot conclude the strong statement that $J^\mu(x)|0\rangle=0$. For if this were true then

$$T^{\mu\nu\alpha}(x,y,z)=\langle 0|T(J^\mu(x)J^\nu(y)J_5^\alpha(z))|0\rangle \tag{4.2}$$

must be purely a seagull, since no matter what the values of x^0, y^0, and z^0 are, there is always an electromagnetic current adjacent to the vacuum. However, a seagull cannot give rise to the anomalous divergence,[1] which survives even in the Gell-Mann–Low scale-invariant limit, since it is mass independent and is not renormalized. *Consequently the equation* $J^\mu(x)|0\rangle=0$ *is false.* Evidently only a weaker statement can be true

$$\langle 0|OJ^\mu(x)|0\rangle=0, \tag{4.3}$$

where O stands for some, *but not all*, operators. In particular, products of electromagnetic currents can comprise O, but O cannot be

$$J_5^\alpha(z)J^\nu(y) \quad \text{or} \quad J^\nu(y)J_5^\alpha(z).$$

A further problem appears if we combine the Federbush-Johnson theorem with the results which we obtained above from Crewther's analysis when fermion creation and annihilation were neglected. As Baker and Johnson[14] have shown, when the coupling g^2 is equal to the value g_0^2 which makes Z_3 finite, and the electron mass is zero, (4.1) holds even in the one-fermion-loop approximation. In particular, g_0^2 is a zero of the function $F(g^2)$ defined in (3.1a) and the four-point function satisfies

$$T^{\mu\nu\lambda\sigma}(x0yz)=\langle 0|T(J^\mu(x)J^\nu(0)J^\lambda(y)J^\sigma(z))|0\rangle^{\text{one-fermion-loop}}_{g^2=g_0^2,\,m=0}=0. \tag{4.4}$$

But now let us take the limit $x\to 0$ in (4.4) and substitute the short-distance expansion of (2.7a). In the one-loop approximation $S_{VV}=F(g_0^2)=0$, so the leading contribution comes from the second term in (2.7a) and is given by

$$T^{\mu\nu\lambda\sigma}(x0yz)\underset{x\sim 0}{\propto}K(g_0^2)\frac{\epsilon^{\mu\nu}{}_{\alpha\beta}x^\beta}{(x^2-i\epsilon)^2}\langle 0|T(J_5^\alpha(0)J^\lambda(y)J^\sigma(z))|0\rangle^{\text{one-fermion-loop}}_{g^2=g_0^2,\,m=0}. \tag{4.5}$$

This is *infinite*, since according to (3.2) the coefficient $K(g_0^2)$ is equal to $F^{-1}(g_0^2)=\infty$, while we have seen that the three-point function appearing in (4.5) cannot vanish. So we have reached the impossible conclusion that $0=\infty$! Evidently, if the theory has an eigenvalue g_0 which makes Z_3 finite, the naive short-distance expansion is invalid *at the eigenvalue*, even though it may be true order-by-order in perturbation theory. In particular, the limiting operations $g\to g_0$ and $x\to 0$ do not commute. One can easily write down simple examples which have this property, e.g.,

$$\frac{F^{-1}(g^2)}{1+f(x)F^{-2}(g^2)}, \tag{4.6}$$

where $f\neq 0$ for $x\neq 0$ but $f(0)=0$. For all nonzero x, (4.6) vanishes as $F(g^2)$ in the limit $g^2\to g_0^2$, but for $x=0$, (4.6) diverges as $F^{-1}(g^2)$ in the same limit. Whether such behavior can actually emerge from field theory, when all the constraints imposed by current conservation and conformal invariance are taken into account, remains an open question, as indeed does the question of whether an eigenvalue g_0^2 exists in the first place.

V. CONCLUSION

We have shown that the coupling-constant-dependent numbers, describing various BJL anomalies, are constrained by the nonrenormalization of the triangle anomaly. Furthermore the axial-vector current continues to behave in a singular fashion even in the finite theory of Gell-Mann and Low. In particular the following three phenomena are incompatible:

(1) The triangle anomaly is unrenormalized.

(2) There is an eigenvalue $g^2=g_0^2$ which makes Z_3 finite.

(3) Naive scale-invariant short-distance expansions involving the axial-vector current are valid at the eigenvalue.

Crewther[5] also applies Wilson's method to anomalies of scale invariance.[1] Unfortunately there does not seem to be a "no renormalization theorem" for these anomalies since *all* regulators vio-

late scale invariance. (Chiral invariance is not violated by boson regulators; these render finite all graphs but the basic fermion triangle.) Therefore results analogous to the above cannot be deduced for scale invariance anomalies.[16]

We have benefited from conversations with R. Crewther, K. Johnson and K. Wilson, which we are happy to acknowledge. SLA and CGC, Jr. wish to acknowledge the hospitality of the National Accelerator Laboratory, where part of this work was done.

APPENDIX

In this Appendix we shall give arguments for our assertion that the vacuum-polarization-free triangle is asymptotically conformal invariant. Our starting point is the Ward identities for scale and conformal invariance. At the naive canonical level, these have the form

$$\int dz \langle 0|T(\Theta(z)\phi^{(1)}(x_1)\cdots\phi^{(n)}(x_n))|0\rangle = i\sum_{i=1}^{n}\left(x_i\cdot\frac{\partial}{\partial x_i}+d_i\right)\langle 0|T(\phi^{(1)}(x_1)\cdots\phi^{(n)}(x_n))|0\rangle, \tag{A1}$$

$$\int dz\, z_\mu \langle 0|T(\Theta(z)\phi^{(1)}(x_1)\cdots\phi^{(n)}(x_n))|0\rangle = i\sum_{i=1}^{n}\left(2x^i_\mu x^i\cdot\frac{\partial}{\partial x^i}-{x_i}^2\frac{\partial}{\partial x_i^\mu}+2x_i^\nu(d_i g_{\mu\nu}+\Sigma^{(i)}_{\mu\nu})\right)\langle 0|T(\phi^{(1)}(x_1)\cdots\phi^{(n)}(x_n))|0\rangle, \tag{A2}$$

where Θ is the trace of the "improved" energy-momentum tensor (hence containing only mass terms and other soft operators), d_i is the canonical dimension of the field $\phi^{(i)}$ and $\Sigma^{(i)}_{\mu\nu}$ is the corresponding intrinsic spin matrix.

The basis for the naive argument for asymptotic conformal invariance is the observation that since Θ must contain explicit factors of mass the left-hand sides of (A1) and (A2) must, on dimensional grounds alone, be less singular at short distances than the corresponding right-hand side.

The work of Zimmermann,[17] Lowenstein,[18] and Schroer[19] indicates that when the unavoidable divergences of perturbation theory are properly taken into account, the above Ward identities are modified by the addition to Θ of operator contributions of dimension four (nonsoft). These new terms have no explicit dimensional factors and need not vanish relative to the right-hand side in the short-distance limit. As a result, asymptotic scale and conformal invariance are not realized in renormalized perturbation theory, except in special cases.

The nonsoft contributions to Θ are associated with the various wave-function and coupling-constant renormalization subtractions needed to make the theory finite. The pieces associated with wave-function renormalization can in fact be absorbed in (A1) and (A2) by replacing the canonical dimensions d_i by coupling-constant-dependent "anomalous dimensions" $\bar{d}_i$. The pieces associated with coupling-constant renormalization are proportional to the various interaction terms in the Lagrangian and simplify only in (A1): The insertion at zero four-momentum of an interaction term is equivalent to differentiation with respect to the corresponding coupling constant.

In the body of the paper we considered a theory of an SU_3 singlet vector-meson coupling via a conserved current to a fermion. The *octet* vector and axial-vector currents in such a theory require no renormalization subtractions, since they cannot be coupled to the singlet vector meson by vacuum polarization bubbles of the type illustrated in Fig. 1. Thus, the $SU_3 \times SU_3$ currents will, according to the preceding paragraph, act like fields with canonical dimensions. The same statement applies to both the electromagnetic and axial-vector currents in quantum electrodynamics with vacuum-polarization insertions omitted. Since the vector meson couples via a conserved current, the usual Ward-identity argument guarantees that coupling-constant infinities arise only from vacuum polarization graphs. If such graphs are excluded – either by fiat, or by looking at a sufficiently low order in perturbation theory – no coupling-constant renormalization is needed, and Θ in (A1) and (A2) may be treated as a soft operator. Further, if we consider a Green's function involving only nonrenormalized currents, so that the relevant dimensions are all canonical, the scale and conformal Ward identities assume their naive form and the argument for asymptotic scale and conformal invariance becomes correct.

Let us apply these remarks to the *VVA* triangle. To $O(g^0)$ (g being the coupling constant of the gluon), we obtain the bare triangle, which is trivially conformally invariant in the short-distance limit. To $O(g^2)$ we obtain the triangle decorated in all possible ways with one gluon. At this level, no vacuum polarization is possible and the above argument indicates that asymptotic conformal invari-

ance still holds. But there is only one possible form for a conformal-invariant VVA amplitude. Therefore, in the short-distance limit, $\Gamma_{VVA} \to (1+cg^2)\Gamma^{(0)}_{VAA}$, where $\Gamma^{(0)}_{VVA}$ stands for the asymptotic limit of the bare triangle. On the other hand, the PCAC (partially conserved axial-vector current) anomaly is determined precisely by the short-distance limit of Γ_{VVA} and is also known to be coupling-constant independent. This is possible only if $C=0$, which is to say that the $O(g^2)$ graphs succeed in being conformal invariant by vanishing. Now consider the $O(g^4)$ contributions to Γ_{VVA}. At this level there *are* vacuum-polarization graphs and the argument for conformal invariance breaks down. Nonetheless scale invariance survives. We argued that when coupling-constant renormalization is needed, (A1) is modified by adding a term

$$\beta(g)\frac{\partial}{\partial g}\langle 0|T(\phi^{(1)}(x_1)\cdots\phi^{(n)}(x_n))|0\rangle$$

to the left-hand side. It turns out that β is $O(g^3)$, so that if we need $\beta(\partial/\partial g)\Gamma_{VVA}$ to $O(g^4)$ it suffices to know Γ_{VVA} to $O(g^2)$. We have just argued that the $O(g^2)$ contribution to Γ_{VVA} vanishes more rapidly in the asymptotic limit than naive power counting would suggest. Therefore, the left-hand side of (A1), computed to $O(g^4)$, still vanishes relative to the right-hand side in the short-distance limit, leading to asymptotic scale invariance. In higher orders, scale invariance presumably breaks down as well.

---

*This work is supported in part through funds provided by the Atomic Energy Commission under Contract AT(11-1)-3069, and by the U. S. Air Force office of Scientific Research under Contract No. F44620-71-C-0108.

†Research sponsored by the National Science Foundation, Grant No. GP-16147 A No. 1.

[1]For a review see S. L. Adler, in *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser, M. Grisaru, and H. Pendleton (MIT Press, Cambridge, Mass., 1970); R. Jackiw, in *Lectures on Current Algebra and Its Applications*, edited by S. Treiman, R. Jackiw, and D. J. Gross (Princeton Univ. Press, Princeton, N. J., 1972).

[2]J. D. Bjorken, Phys. Rev. 148, 1467 (1966); K. Johnson and F. E. Low, Progr. Theoret. Phys. (Kyoto) Suppl. 37-38, 74 (1966).

[3]C. G. Callan, Jr. and D. J. Gross, Phys. Rev. Letters 22, 156 (1969).

[4]D. G. Sutherland, Nucl. Phys. B2, 433 (1967); M. Veltman, Proc. Roy. Soc. (London) A301, 107 (1967).

[5]R. J. Crewther, Phys. Rev. Letters 28, 1421 (1972).

[6]K. G. Wilson, Phys. Rev. 179, 1499 (1969).

[7]M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954). For a recent review see S. L. Adler, Phys. Rev. D 5, 3021 (1972).

[8]E. J. Schreier, Phys. Rev. D 3, 982 (1971). Schreier's result may be readily understood in the 6-dimensional conformal-covariant formalism of Dirac [P. A. M. Dirac, Ann. Math. 37, 429 (1936); we follow the notation of G. Mack and A. Salam, Ann. Phys. (N.Y.) 53, 174 (1969)], where it is the statement that the only 3-index pseudotensor which can be constructed from three coordinates η_1, η_2, η_3, and is homogeneous of degree -3 in each coordinate, is $\epsilon_{ABCDEF}\eta_1^D\eta_2^E\eta_3^F/(\eta_1\cdot\eta_2\eta_2\cdot\eta_3\eta_3\cdot\eta_1)^2$.

[9]Our conventions are the following: $\{\gamma^\mu,\gamma^\nu\}=2g^{\mu\nu}$; $g^{\mu\nu}=0$, $\mu\neq\nu$, $g^{00}=-g^{11}=-g^{22}=-g^{33}=1$; $\gamma^5=\gamma^0\gamma^1\gamma^2\gamma^3$. We assume SU(3) symmetry realized by triplet quarks, hence the occurrence of $d_{abc}=\frac{1}{4}\,\mathrm{Tr}\,\{\lambda_a,\lambda_b\}\lambda_c$.

[10]Henceforth we take $N=N'$; $S_{VV}=S_{AA}=S$; $K_{VV}=K_{VA}=K_{AA}$.

[11]For details see S. L. Adler, Ref. 1.

[12]R. Jost and J. Luttinger, Helv. Phys. Acta. 23, 201 (1950).

[13]M. Baker and K. Johnson, Phys. Rev. D 3, 2541 (1971).

[14]M. Baker and K. Johnson, Phys. Rev. 183, 1292 (1969).

[15]It has already been shown by M. Baker and K. Johnson, Phys. Rev. D 3, 2516 (1971), that the commutator of fermion fields is finite, but anomalous in the finite theory.

[16]M. Chanowitz and J. Ellis [Phys. Letters 40B, 397 (1972)] also have studied scale-invariance anomalies by using a momentum-space version of Crewther's method.

[17]W. Zimmermann, in *Lectures on Elementary Particles and Quantum Field Theory*, Ref. 1.

[18]J. H. Lowenstein, Phys. Rev. D 4, 2281 (1971).

[19]B. Schroer, Lett. Nuovo Cimento 2, 867 (1971).

PHYSICAL REVIEW D VOLUME 6, NUMBER 12 15 DECEMBER 1972

Massless, Euclidean Quantum Electrodynamics on the 5-Dimensional Unit Hypersphere

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540
and National Accelerator Laboratory, Batavia, Illinois 60510*
(Received 21 August 1972)

We show that the Feynman rules for vacuum-polarization calculations and the equations of motion in massless, Euclidean quantum electrodynamics can be transcribed, by means of a stereographic mapping, to the surface of the 5-dimensional unit hypersphere. The resulting formalism is closely related to the Feynman rules, which we also develop, for massless electrodynamics in the conformally covariant O(5,1) language. The hyperspherical formulation has a number of apparent advantages over conventional Feynman rules in Euclidean space: It is manifestly infrared-finite, and it may permit the development of approximation methods based on a semiclassical approximation for angular momenta on the hypersphere. The finite-electron-mass, Minkowski-space generalization of our results gives a simple formulation of electrodynamics in (4,1) de Sitter space.

I. INTRODUCTION

Conformal invariance in quantum field theory has attracted renewed interest recently, because of its connection with problems of asymptotic high-energy behavior.[1] Important results on leading light-cone singularities, for example, have been obtained by the use of conformal invariance.[2] Another question to which conformal invariance is relevant is the study of eigenvalue conditions imposed by requiring renormalization constants to be finite.[3] To see this, let us consider the single-fermion-loop vacuum-polarization diagrams in spin-$\frac{1}{2}$ quantum electrodynamics, illustrated in Fig. 1. If we work in coordinate space with separated points x, x' we can freely pass to the zero fermion mass, or conformal limit. In this limit, however, the structure of the vacuum polarization is unique,[2] and hence the sum of diagrams in Fig. 1(a) must be proportional to the lowest-order vacuum-polarization tensor in Fig. 1(b),

$$\pi_{\mu\nu}(x, x'; \alpha) = -3\pi F^{[1]}(\alpha)\pi^{(0)}_{\mu\nu}(x, x') . \qquad (1)$$

When Eq. (1) is Fourier-transformed to momentum space, using current conservation in the usual fashion to eliminate the quadratic divergence, the function $F^{[1]}(\alpha)$ appears as the coefficient of the logarithmically divergent term. Requiring the photon wave-function renormalization Z_3 to be finite then imposes the eigenvalue condition $F^{[1]}(\alpha) = 0$.[4]

Our aim in the present paper is to study reformulations of massless electrodynamics which are made possible by its invariance under conformal transformations, with the goal of developing methods which may allow one to calculate or approximate the function $F^{[1]}$ appearing in Eq. (1). Because the singularity structure in x and x' is not of interest (it is just that of the lowest-order vacuum polarization), we make the Dyson-Wick rotation to a Euclidean metric at the outset. Thus we deal with massless, Euclidean quantum electrodynamics. Our principal result is that the Feynman rules for vacuum-polarization calculations and the equations of motion in this theory can be simply rewritten in terms of equivalent rules and equations of motion on the surface of the 5-dimensional unit hypersphere. In Sec. II we state the 5-dimensional rules and verify by explicit transformation that they are equivalent to the usual rules in Euclidean coordinate space (x space). We also construct and verify a 5-dimensional formulation of the Maxwell equations and the equation of current conservation, and discuss the physical meaning of rotations and inversions on the hypersphere. In Sec. III we discuss massless, Euclidean quantum electrodynamics in the manifestly conformal-covariant O(5,1) language. We develop the Feynman rules in this formalism, explore some of their peculiar features, and show that they are related by a simple projective transformation to the rules on the 5-dimensional hypersphere. In Sec. IV we discuss possible generalizations and applications of our results. We point out that the finite-electron-mass, Minkowski-space extension of our hyperspherical results gives a simple formulation of electrodynamics in (4,1) de Sitter space. The electron wave equation which we use is just the de Sitter-space equation originally proposed by Dirac,[5] but our treatment of the Maxwell equations is an improvement over that of Dirac, and does not require the imposition of homogeneity condi-

tions. There are a number of possible calculational advantages of the hyperspherical formulation of electrodynamics over the usual Feynman rules in Euclidean space. First, because the surface of the hypersphere is a bounded domain, the calculation of vacuum-polarization diagrams in the 5-dimensional formalism is manifestly infrared-finite. Second, because the wave operators on the hypersphere are constructed from angular momentum operators, there appears to be the possibility of making semiclassical approximations when virtual angular momentum quantum numbers are large compared to unity. This contrasts sharply with the situation in Euclidean space, where there is no natural distance or momentum scale which distinguishes regions where one can approximate the wave operator.

II. 5-DIMENSIONAL FORMALISM

In this section we set out the 5-dimensional formalism and verify, by explicit transformation, its equivalence to the usual rules in x space. Secs. IIA–IIC contain a summary of the 5-dimensional Feynman rules and equations of motion, while in Secs. IID and IIE we discuss the transformation to x space and the interpretation of symmetries on the hypersphere.

A. Summary of Feynman Rules on the Hypersphere

In writing down the 5-dimensional rules and comparing them with their Euclidean counterparts, we adhere to the following conventions and notation.[6] Five-dimensional unit vectors are denoted by $\eta_1, \eta_2, \ldots$; 5-dimensional vector indices are indicated by lower case italic letters $a, b, \ldots$ which take the values $1, \ldots, 5$, and the 5-dimensional metric is the Euclidean metric δ_{ab}. Similarly, ordinary 4-dimensional vectors are denoted by $x_1, x_2, \ldots$ with vector indices $\mu, \nu, \ldots$ taking the values $1, \ldots, 4$ and with a 4-dimensional Euclidean metric $\delta_{\mu\nu}$. The usual 4×4 Dirac γ matrices are taken to satisfy a Euclidean Clifford algebra

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu} \tag{2}$$

and are all Hermitian; explicit representations for these matrices are well known. In writing the 5-dimensional rules we need, instead of the γ's, a set of five Hermitian 8×8 matrices α_a satisfying the Clifford algebra

$$\{\alpha_a, \alpha_b\} = 2\delta_{ab}\,. \tag{3}$$

In terms of the γ matrices and the Pauli spin matrices $\tau_{1,2,3}$, an explicit representation of the α matrices is

$$\alpha_\mu = \gamma_\mu \tau_1, \quad \alpha_5 = \tau_3\,. \tag{4}$$

Since the matrix

$$\alpha_6 = \tau_2 \tag{5}$$

satisfies $\alpha_6{}^2 = 1$ and anticommutes with the α_a, the trace of an odd number of α matrices vanishes. Physical quantities such as the electromagnetic current, vector potential, etc., will be denoted by capital letters $(J_a, A_a, \ldots)$ in 5-dimensional space and by lower case letters $(j_\mu, a_\mu, \ldots)$ in Euclidean space. We let $\int d^4x = \int dx_1 dx_2 dx_3 dx_4$ denote the integration of x over Euclidean space and we similarly let $\int d\Omega_\eta$ denote the integration of η over the surface of the 5-dimensional hypersphere. Finally, we use tr_4 and tr_8 to denote, respectively, the trace over the γ matrices and the α matrices.

The connection between the 5-dimensional co-

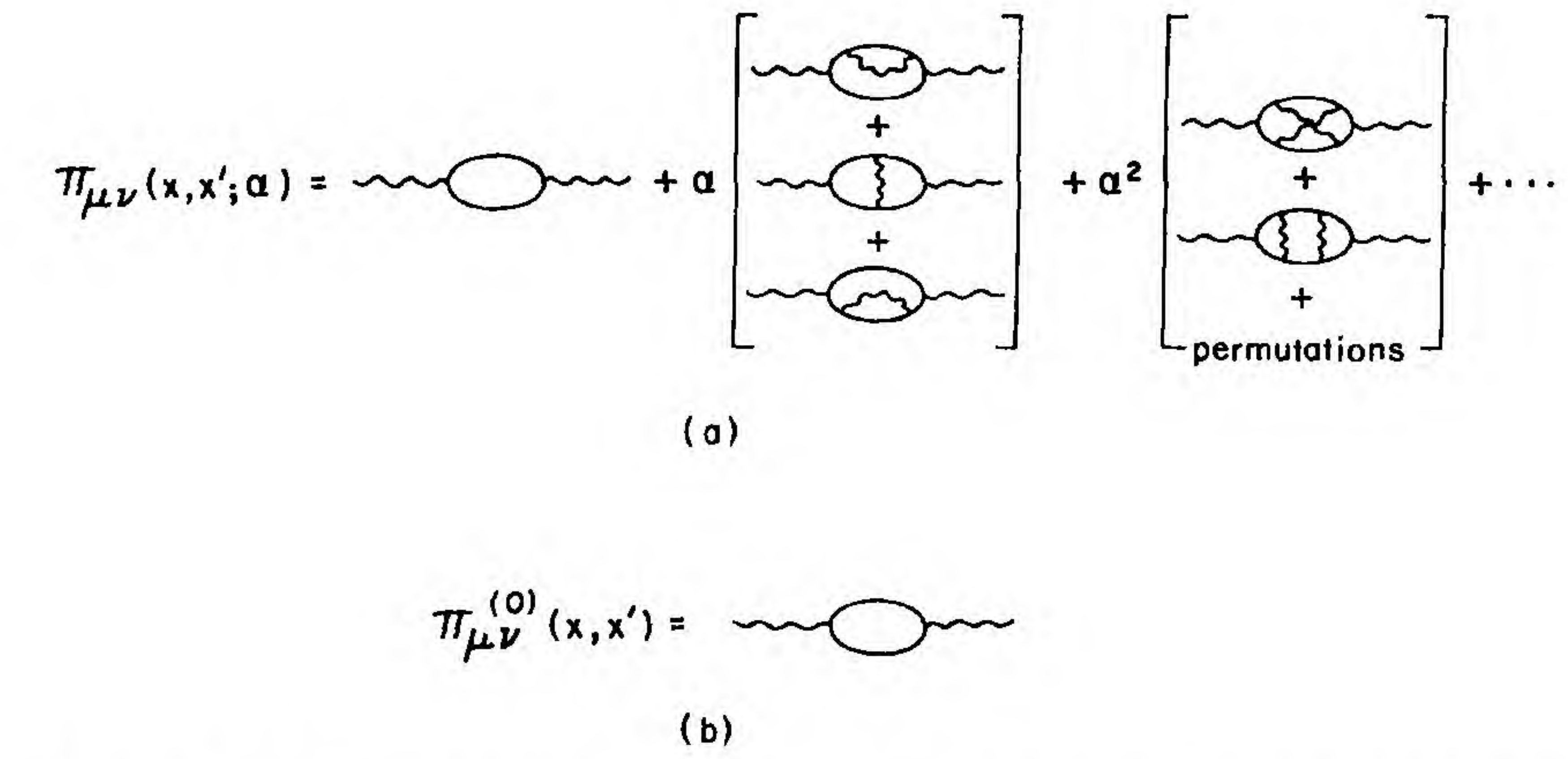


FIG. 1. (a) Single-fermion-loop vacuum-polarization diagrams in spin-½ electrodynamics. (b) Lowest-order vacuum-polarization diagram.

ordinate η describing a space-time point and its Euclidean equivalent x is given by the stereographic mapping[7]

$$x_\mu = \kappa^{-1}\eta_\mu, \quad \kappa = 1+\eta_5, \tag{6a}$$

with the inverse transformation

$$\eta_\mu = \frac{2x_\mu}{1+x^2}, \quad \eta_5 = \frac{1-x^2}{1+x^2}. \tag{6b}$$

The 5-dimensional electromagnetic current J_a, which satisfies the constraint equation

$$\eta \cdot J = \eta_a J_a = 0, \tag{7}$$

is mapped into the usual electromagnetic current j_μ by

$$\kappa^{-3} j_\mu = J_\mu - x_\mu J_5, \tag{8}$$

with the inverse transformation

$$\begin{aligned} J_\mu &= \kappa^{-3} j_\mu - \kappa^{-2} x_\mu x\cdot j, \\ J_5 &= -\kappa^{-2} x\cdot j \,. \end{aligned} \tag{9}$$

We can now state the 5-dimensional Feynman rules with, for comparison, their Euclidean counterparts. These are given in Table I. The equivalence of the two sets of Feynman rules for vacuum polarization (closed-fermion-loop) calculations is demonstrated explicitly in Sec. IID below.

B. Photon Propagator Equation and Maxwell Equations

To write the wave equation satisfied by the photon propagator on the hypersphere we introduce the (anti-Hermitian) angular momentum operator

$$L_{ab} = \eta_a \frac{\partial}{\partial \eta_b} - \eta_b \frac{\partial}{\partial \eta_a}. \tag{10}$$

When several coordinates $\eta_1, \eta_2, \ldots$ are present we denote the angular momentum acting at η_1 by L_1, so $(L_1)_{ab} = (\eta_1)_a \partial/\partial(\eta_1)_b - (\eta_1)_b\partial/\partial(\eta_1)_a$, etc. In this notation the photon propagator equation takes the form

$$(L_1^2 - 4)\frac{1}{(\eta_1-\eta_2)^2} = -8\pi^2 \delta_S(\eta_1-\eta_2), \tag{11}$$

where δ_S is the hyperspherical δ function satisfying

$$\int d\Omega_{\eta_1} f(\eta_1)\delta_S(\eta_1-\eta_2) = f(\eta_2) \tag{12}$$

for arbitrary f. The constant multiplying the δ_S function in Eq. (11) can be verified by integrating Eq. (11) over the hypersphere:

$$\begin{aligned} \int d\Omega_1 (L_1^2-4)\frac{1}{(\eta_1-\eta_2)^2} &= -4\int d\Omega_1 \frac{1}{(\eta_1-\eta_2)^2} \\ &= -4\left[\frac{\int_{-1}^1 d\mu(1-\mu^2)[1/2(1-\mu)]}{\int_{-1}^1 d\mu(1-\mu^2)}\right]\int d\Omega_1 \\ &= -4\left(\frac{1}{4/3}\right)\frac{8\pi^2}{3} \\ &= -8\pi^2, \end{aligned} \tag{13}$$

where we have written $\mu = \eta_1 \cdot \eta_2$ and used the fact that

$$\int d\Omega_1 = \int_{-1}^{1} d\mu (1-\mu^2) \times \text{azimuthal integrations}. \tag{14}$$

Equation (11) can also be verified from the expansion of $(1-\mu)^{-1}$ in terms of Gegenbauer polynomials $C_n^{3/2}(\mu)$,[8]

TABLE I. The 5-dimensional Feynman rules and their Euclidean counterparts.

	5-dimensional	Euclidean
Electron propagator	$\frac{-i}{\pi^2}\frac{\frac{1}{2}(\alpha\cdot\eta_1-1)\frac{1}{2}(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4}$	$\frac{-i}{2\pi^2}\frac{\gamma\cdot(x_1-x_2)}{(x_1-x_2)^4}$
Photon propagator	$\frac{1}{4\pi^2}\frac{\delta_{ab}}{(\eta_1-\eta_2)^2}$	$\frac{1}{4\pi^2}\frac{\delta_{\mu\nu}}{(x_1-x_2)^2}$ + gauge terms
Electron-photon vertex [a]	$ie\alpha_a \equiv \frac{1}{2}ie[\alpha\cdot\eta_1, \alpha_a]$	$ie\gamma_\mu$
Each closed fermion loop	$-\mathrm{tr}_8$	$-\mathrm{tr}_4$
Each virtual coordinate integration	$\int d\Omega_\eta$	$\int d^4x$

[a] The two indicated forms of 5-dimensional electron-photon vertex are equal when sandwiched between electron propagators.

$$\frac{1}{1-\mu}=\sum_{n=0}^{\infty}\frac{(2n+3)C_n^{3/2}(\mu)}{(n+1)(n+2)} . \tag{15}$$

Using the relation

$$\sum_m Y_{nm}(\eta_1)Y_{nm}^*(\eta_2)=\frac{2n+3}{8\pi^2}C_n^{3/2}(\eta_1\cdot\eta_2), \tag{16}$$

where the $Y_{nm}(\eta)$ are orthonormalized hyperspherical harmonics, Eq. (15) becomes

$$\frac{1}{(\eta_1-\eta_2)^2}=4\pi^2\sum_{n=0}^{\infty}\sum_m\frac{Y_{nm}(\eta_1)Y_{nm}^*(\eta_2)}{(n+1)(n+2)} . \tag{17}$$

Then, using the differential equation for the hyperspherical harmonics

$$L_1^2Y_{nm}(\eta_1)=-2n(n+3)Y_{nm}(\eta_1), \tag{18}$$

we find from Eq. (17) that

$$\begin{aligned}(L_1^2-4)\frac{1}{(\eta_1-\eta_2)^2}&=-8\pi^2\sum_{n=0}^{\infty}\sum_m Y_{nm}(\eta_1)Y_{nm}^*(\eta_2)\\&=-8\pi^2\delta_S(\eta_1-\eta_2),\end{aligned} \tag{19}$$

in agreement with Eq. (11).

To write the Maxwell equations on the hypersphere we introduce the electromagnetic potential 5-vector A_a, which satisfies the constraint

$$\eta\cdot A=0 \tag{20}$$

and is related to the electromagnetic potential a_μ in Euclidean space by

$$\kappa^{-1}a_\mu=A_\mu-x_\mu A_5 . \tag{21}$$

The electromagnetic field strength is described by the totally antisymmetric rank-three tensor

$$F_{abc}=L_{ab}A_c+L_{bc}A_a+L_{ca}A_b , \tag{22}$$

which is dual to the antisymmetric rank-two tensor

$$\begin{aligned}\hat F_{ab}&=\tfrac{1}{6}\epsilon_{abcde}F_{cde}\\&=\epsilon_{abcde}\eta_c\frac{\partial}{\partial\eta_d}A_e .\end{aligned} \tag{23}$$

The usual dual tensor $\hat f_{\mu\nu}$ in Euclidean space is related to $\hat F_{ab}$ by

$$\kappa^{-2}\hat f_{\mu\nu}=\hat F_{\mu\nu}-x_\mu\hat F_{5\nu}-x_\nu\hat F_{\mu5} . \tag{24}$$

In terms of the tensors F_{abc} and $\hat F_{ab}$ the Maxwell equations become

$$L_{ab}F_{abc}=2eJ_c , \tag{25a}$$

$$L_{ab}\hat F_{bc}=\hat F_{ac} , \tag{25b}$$

with J_c the electromagnetic current, which satisfies the conservation equation[8]

$$L_{ab}J_b=J_a . \tag{26}$$

An explicit demonstration that Eqs. (25) and (26) indeed do correspond to the Maxwell and current-conservation equations in x space will be given below in Sec. IID.

When Eqs. (22) and (23) are used to express the Maxwell equations in terms of the potential A, Eq. (25b) is trivially satisfied, while Eq. (25a) becomes

$$P_{ca}A_a=2eJ_c , \tag{27}$$

with P_{ca} the wave operator

$$P_{ca}=2L_{cb}L_{ba}-6L_{ca}+L^2\delta_{ca} . \tag{28}$$

Using the angular momentum commutation relations it is straightforward to verify that P_{ca} has the following properties:

$$\begin{aligned}&L_{bc}P_{ca}=P_{bc}L_{ca}=P_{ba} ,\\&\eta_bP_{ba}=P_{ba}\eta_a=0 ,\end{aligned} \tag{29}$$

which guarantee the consistency of Eq. (27) with the constraints on J given by Eq. (7) and Eq. (26). Equation (27) can be further simplified if the potential A_a is chosen to satisfy the condition

$$L_{ab}A_b=A_a , \tag{30}$$

which is the hyperspherical analog of the Lorentz condition. When acting on potentials which obey Eq. (30) the operator P_{ca} becomes simply $(L^2-4)\delta_{ca}$. Hence the wave equation becomes

$$(L^2-4)A_c=2eJ_c , \tag{31}$$

and as expected involves the same wave operator as appears in the photon propagator equation, Eq. (11).

C. Electron Propagator Equation and Field Equation

To write the electron propagator equation we introduce the matrix γ_{ab} defined by

$$\gamma_{ab}=\tfrac{1}{4}i[\alpha_a,\alpha_b] . \tag{32}$$

Using the abbreviation $\gamma_{ab}L_{ab}=\gamma\cdot L$, we find that the electron propagator obeys the wave equation

$$\begin{aligned}(i\gamma\cdot\vec L_1+2)\frac{(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4}&=-2\pi^2\delta_S(\eta_1-\eta_2)(\alpha\cdot\eta_2+1),\\ \frac{(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4}(i\gamma\cdot\overleftarrow{L}_2-2)&=-2\pi^2\delta_S(\eta_1-\eta_2)(\alpha\cdot\eta_1-1),\end{aligned} \tag{33}$$

where the coefficient of the δ_S function in Eq. (33) is obtained by averaging over the hypersphere, as in Eq. (13). An alternative method for obtaining Eq. (33) is to use the following relation between the electron and photon propagators,

$$\frac{(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4} = -\tfrac{1}{2}(i\gamma\cdot\vec{L}_1+1)\frac{1}{(\eta_1-\eta_2)^2}(\alpha\cdot\eta_2+1). \quad (34)$$

Applying the wave operator $i\gamma\cdot\vec{L}_1+2$ to Eq. (34) and using the identity

$$(i\gamma\cdot L_1+2)(i\gamma\cdot L_1+1)=-\tfrac{1}{2}(L_1^2-4), \quad (35)$$

we find

$$\begin{aligned}(i\gamma\cdot\vec{L}_1+2)\frac{(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4} &= \tfrac{1}{4}(L_1^2-4)\frac{1}{(\eta_1-\eta_2)^2}(\alpha\cdot\eta_2+1)\\ &= -2\pi^2\delta_S(\eta_1-\eta_2)(\alpha\cdot\eta_2+1),\end{aligned} \quad (36)$$

where in the last step we have used the photon propagator equation, Eq. (11).

The matrix γ_{ab} is a generalized spin operator for the electron. Writing

$$\begin{aligned}S_{ab}&=-i\gamma_{ab}\\ &=\tfrac{1}{4}[\alpha_a,\alpha_b],\end{aligned} \quad (37)$$

we find that S and L satisfy identical commutation relations,

$$\begin{aligned}[S_{ab},S_{cd}]&=\delta_{ac}S_{db}-\delta_{ad}S_{cb}+\delta_{bc}S_{ad}-\delta_{bd}S_{ac},\\ [L_{ab},L_{cd}]&=\delta_{ac}L_{db}-\delta_{ad}L_{cb}+\delta_{bc}L_{ad}-\delta_{bd}L_{ac},\end{aligned} \quad (38)$$

and that

$$\begin{aligned}S^2&=-5,\\ (L\cdot S)^2&=3L\cdot S-\tfrac{1}{2}L^2.\end{aligned} \quad (39)$$

The second relation in Eq. (39) leads immediately to the identity in Eq. (35).

Finally, in terms of an 8-component electron spinor χ the electron wave equation takes the form

$$\left\{i\gamma_{ab}\left[\eta_a\left(\frac{\partial}{\partial\eta_b}-ieA_b(\eta)\right)-\eta_b\left(\frac{\partial}{\partial\eta_a}-ieA_a(\eta)\right)\right]+2\right\}\chi=0, \quad (40)$$

with the adjoint equation

$$\bar{\chi}\left\{i\gamma_{ab}\left[\eta_a\left(\frac{\overleftarrow{\partial}}{\partial\eta_b}+ieA_b(\eta)\right)-\eta_b\left(\frac{\overleftarrow{\partial}}{\partial\eta_a}+ieA_a(\eta)\right)\right]-2\right\}=0,\qquad \bar{\chi}=\chi^\dagger. \quad (41)$$

The electromagnetic current J_c which appears in Eq. (31) is given by

$$J_c(\eta)=-\tfrac{1}{2}i\bar{\chi}[\alpha\cdot\eta,\alpha_c]\chi. \quad (42)$$

Using the relation

$$\left(\eta_a\frac{\partial}{\partial\eta_c}-\eta_c\frac{\partial}{\partial\eta_a}\right)[\alpha\cdot\eta,\alpha_c]=[\alpha\cdot\eta,\alpha_a]-2i\eta_a\gamma\cdot L \quad (43)$$

and Eqs. (40) and (41), we see that the current J_c satisfies the current conservation condition of Eq. (26). The constraint imposed by Eq. (7) is also obviously satisfied.

D. Transformation from the Hypersphere to x Space

We give in this section the explicit transformations which map the hyperspherical Feynman rules and equations of motion into the corresponding rules and equations of motion in x space. We begin with the Feynman rules of Table I and consider first a closed fermion loop coupling to $2n$ photons, given by

$$\begin{aligned}\frac{1}{2n}\sum_{\text{permutations of }1,\ldots,2n} -\mathrm{tr}_8\Bigg[&\frac{-i}{\pi^2}\frac{\frac{1}{2}(\alpha\cdot\eta_1-1)\frac{1}{2}(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4}\,ie\alpha_{a_2}\,\frac{-i}{\pi^2}\frac{\frac{1}{2}(\alpha\cdot\eta_2-1)\frac{1}{2}(\alpha\cdot\eta_3+1)}{(\eta_2-\eta_3)^4}\cdots\\ &\times\frac{-i}{\pi^2}\frac{\frac{1}{2}(\alpha\cdot\eta_{2n}-1)\frac{1}{2}(\alpha\cdot\eta_1+1)}{(\eta_{2n}-\eta_1)^4}\,ie\alpha_{a_1}\Bigg].\end{aligned} \quad (44)$$

In order to obtain the corresponding $2n$-point function in x space, we must transform each of the $2n$ current indices according to the recipe of Eq. (8a), which means that we effectively make the replacement

$$\alpha_{a_i}\to\kappa_i^3(\alpha_{\mu_i}-x_{\mu_i}\alpha_5) \quad (45)$$

for each vertex α_{a_i}. After this replacement has been made, we must then find that a purely *algebraic* rearrangement of factors gives the $2n$-point function computed from x-space Feynman rules. For the denominator in Eq. (44) the rearrangement is trivial, since substitution of Eq. (6) shows that

$$(\eta_i-\eta_{i+1})^2=\kappa_i\kappa_{i+1}(x_i-x_{i+1})^2. \quad (46)$$

To rearrange the numerator we exploit the fact that the factors $\alpha\cdot\eta\pm1$ appearing in each propa-

gator are projection operators, allowing us to rewrite the numerator of the general propagator according to

$$\tfrac{1}{2}(\alpha\cdot\eta_i-1)\tfrac{1}{2}(\alpha\cdot\eta_{i+1}+1)$$
$$=\tfrac{1}{2}(\alpha\cdot\eta_i-1)\tfrac{1}{2}\alpha\cdot(\eta_{i+1}-\eta_i)\tfrac{1}{2}(\alpha\cdot\eta_{i+1}+1). \quad (47)$$

We next introduce the matrix $O(x)$ given by

$$O(x)=\frac{1+\alpha_5\alpha\cdot x}{(1+x^2)^{1/2}},$$
$$O(x)^{-1}=\frac{1-\alpha_5\alpha\cdot x}{(1+x^2)^{1/2}}, \quad (48)$$

where $\alpha\cdot x$ denotes the 4-dimensional scalar product $\alpha_\mu x_\mu$. Some straightforward algebra then shows that

$$O(x_i)\tfrac{1}{2}\alpha\cdot(\eta_i-\eta_{i+1})O(x_{i+1})^{-1}$$
$$=\frac{\alpha\cdot(x_i-x_{i+1})}{[(1+x_i^2)(1+x_{i+1}^2)]^{1/2}}$$
$$=(\tfrac{1}{2}\kappa_i\tfrac{1}{2}\kappa_{i+1})^{1/2}\alpha\cdot(x_i-x_{i+1}) \quad (49)$$

and that

$$O(x)\tfrac{1}{2}(1+\alpha\cdot\eta)(\alpha_\mu-x_\mu\alpha_5)\tfrac{1}{2}(\alpha\cdot\eta-1)O(x)^{-1}$$
$$=\tfrac{1}{2}(1+\alpha_5)\alpha_\mu\tfrac{1}{2}(\alpha_5-1). \quad (50)$$

Substituting Eq. (4) for the α matrices, we can pull all factors $\frac{1}{2}(\alpha_5\pm1)=\frac{1}{2}(\tau_3\pm1)$ to the left, where they combine to give a single factor $\frac{1}{2}(\tau_3+1)$. The factors τ_1 appearing in the matrices α_μ then cancel in pairs $(\tau_1^2=1)$, leaving

$$-\mathrm{tr}_8[\tfrac{1}{2}(\tau_3+1)X\{\gamma\}]=-\mathrm{tr}_4[X\{\gamma\}], \quad (51)$$

where $X\{\gamma\}$ contains γ matrices only. The factor κ_i^3 appearing in Eq. (45) precisely cancels the factor $(\kappa_i^{1/2}/\kappa_i^2)^2$ arising from the substitution of Eq. (49) and Eq. (46) into Eq. (44). Thus, we have shown that when the replacements of Eq. (45) are made, Eq. (44) can be algebraically rearranged to the form

$$\frac{1}{2n}\sum_{\substack{\text{permutations}\\ \text{of } 1,\dots,2n}}-\mathrm{tr}_4\left[\frac{-i}{2\pi^2}\frac{\gamma\cdot(x_1-x_2)}{(x_1-x_2)^4}ie\gamma_{\mu_2}\frac{-i}{2\pi^2}\frac{\gamma\cdot(x_2-x_3)}{(x_2-x_3)^4}\cdots\frac{-i}{2\pi^2}\frac{\gamma\cdot(x_{2n}-x_1)}{(x_{2n}-x_1)^4}ie\gamma_{\mu_1}\right], \quad (52)$$

which is just the $2n$-point function calculated according to the Euclidean x-space Feynman rules.

The next step is to verify that the hyperspherical rule,

$$\int d\Omega_{\eta_1}d\Omega_{\eta_2}J_{a_1}(\eta_1)J_{a_2}(\eta_2)\frac{1}{4\pi^2}\frac{\delta_{a_1a_2}}{(\eta_1-\eta_2)^2}, \quad (53)$$

correctly describes the propagation of a virtual photon from η_1 to η_2. The use of the current J in Eq. (53) is of course just a convenient shorthand for describing the $2n$-point functions from which the photon is emitted (absorbed), with all variables other than those referring to the virtual photon in question suppressed. Calculating the Jacobian of the transformation of Eq. (6) by use of 5-dimensional spherical coordinates gives

$$d\Omega_\eta=\kappa^4d^4x. \quad (54)$$

Substituting Eq. (54) into Eq. (53), using Eq. (46) to rewrite the denominator of the photon propagator and Eq. (9) to reexpress the current J_a in terms of the combination of components $j_\mu=\kappa^3(J_\mu-x_\mu J_5)$ which is relevant to x space, we find that the factors κ precisely cancel, leaving

$$\int d^4x_1d^4x_2j_{\mu_1}(x_1)j_{\mu_2}(x_2)\frac{1}{4\pi^2}\Delta_{\mu_1\mu_2}(x_1,x_2), \quad (55)$$

with

$$\Delta_{\mu_1\mu_2}(x_1,x_2)=\frac{1}{(x_1-x_2)^2}\left[\delta_{\mu_1\mu_2}-\frac{2(x_1)_{\mu_1}(x_1)_{\mu_2}}{1+x_1^2}-\frac{2(x_2)_{\mu_1}(x_2)_{\mu_2}}{1+x_2^2}+\frac{4(1+x_1\cdot x_2)(x_1)_{\mu_1}(x_2)_{\mu_2}}{(1+x_1^2)(1+x_2^2)}\right] \quad (56a)$$

$$=\frac{\delta_{\mu_1\mu_2}}{(x_1-x_2)^2}+\frac{\partial}{\partial(x_2)_{\mu_2}}\left[\ln(x_1-x_2)^2\frac{1}{2}\frac{\partial}{\partial(x_1)_{\mu_1}}\ln(1+x_1^2)\right]$$
$$+\frac{\partial}{\partial(x_1)_{\mu_1}}\left[\ln(x_1-x_2)^2\frac{1}{2}\frac{\partial}{\partial(x_2)_{\mu_2}}\ln(1+x_2^2)\right]-\frac{\partial}{\partial(x_1)_{\mu_1}}\frac{\partial}{\partial(x_2)_{\mu_2}}[\tfrac{1}{2}\ln(1+x_1^2)\ln(1+x_2^2)]. \quad (56b)$$

In Eq. (56a) we give the form of the x-space photon propagator $\Delta_{\mu_1\mu_2}(x_1, x_2)$ which emerges directly from the substitution of Eq. (9) into Eq. (53); in Eq. (56b) we show that Δ can be rewritten as the usual Feynman propagator plus total derivative terms (gauge terms), which make no contribution to Eq. (55) because of electromagnetic current conservation. Hence Eq. (53) is completely equivalent to the usual x-space Feynman rules for propagating a virtual photon. In Sec. II E we will show that the special significance of the gauge terms in Eq. (56a) is that they give Δ simple transformation properties under coordinate inversion.

To transform the photon and electron equations of motion and the current conservation equation to x space, we rewrite the differential operators $\partial/\partial\eta_a$ and L_{ab} in terms of x derivatives according to

$$\frac{\partial}{\partial\eta_a} = \kappa^{-1}(\delta_{a\mu} - \delta_{a5}x_\mu)\frac{\partial}{\partial x_\mu} + \text{terms proportional to } \eta_a, \tag{57}$$

$$L_{\mu\nu} = x_\mu \frac{\partial}{\partial x_\nu} - x_\nu \frac{\partial}{\partial x_\mu},$$
$$L_{5\mu} = x_\mu x\cdot\frac{\partial}{\partial x} + (1-\kappa^{-1})\frac{\partial}{\partial x_\mu}, \tag{58a}$$

and use the following equation [obtained from Eq. (6)] to differentiate κ,

$$\frac{\partial\kappa}{\partial x_\mu} = -\kappa^2 x_\mu\,. \tag{58b}$$

Applying Eqs. (58) to Eqs. (21)–(24) we find that Eq. (24) implies the usual connection between the x-space field strength $\hat{f}_{\mu\nu}$ and potential a_μ,

$$f_{\lambda\sigma} \equiv \frac{\partial}{\partial x_\lambda}a_\sigma - \frac{\partial}{\partial x_\sigma}a_\lambda$$
$$= \tfrac{1}{2}\epsilon_{\lambda\sigma\mu\nu}\hat{f}_{\mu\nu}\,. \tag{59}$$

Similarly, applying Eqs. (58) to Eqs. (25) and (26), we find (after considerable algebra) that these equations reduce to the usual Maxwell and current conservation equations in x space:

$$\text{Eq. (25a)} \Rightarrow \partial_\mu f_{\mu\nu} = ej_\nu\,, \tag{60a}$$

$$\text{Eq. (25b)} \Rightarrow \partial_\mu \hat{f}_{\mu\nu} = 0\,, \tag{60b}$$

$$\text{Eq. (26)} \Rightarrow \partial_\mu j_\mu = 0\,. \tag{61}$$

To transform the electron wave equation [Eq. (40)] and the expression for the electromagnetic current [Eq. (42)] to x space, we first note that the Pauli matrix τ_2 commutes with the wave operator in Eq. (40), and hence the 4-component spinors

$$\chi_\pm = \tfrac{1}{2}(1\pm\tau_2)\chi \tag{62}$$

also satisfy Eq. (40). Defining x-space 4-component spinors $\psi_\pm$ by

$$\psi_\pm = \kappa^{3/2}O(x)\chi_\pm\,, \tag{63}$$

we find that the projection of Eq. (42) into x space takes the form

$$j_\mu(x) = -\tfrac{1}{2}i\kappa^3\bar{\chi}[\alpha\cdot\eta,\ \alpha_\mu - x_\mu\alpha_6]\chi$$
$$= j_\mu^+ - j_\mu^-\,, \tag{64}$$

with

$$j_\mu^\pm = \bar{\psi}_\pm\gamma_\mu\psi_\pm\,,\quad \bar{\psi}_\pm = \psi_\pm^\dagger\,. \tag{65}$$

Since a direct (and again somewhat lengthy) calculation shows that the Dirac wave operator obeys the transformation

$$\kappa^{3/2}O(x)\left\{i\gamma_{ab}\left[\eta_a\left(\frac{\partial}{\partial\eta_b} - ieA_b(\eta)\right) - \eta_b\left(\frac{\partial}{\partial\eta_a} - ieA_a(\eta)\right)\right] + 2\right\}\kappa^{-3/2}O(x)^{-1} = -i\tau_2\kappa^{-1}\gamma\cdot\left(\frac{\partial}{\partial x} - iea\right), \tag{66}$$

the x-space spinors $\psi_\pm$ satisfy the usual mass-zero Dirac equation

$$\gamma\cdot\left(\frac{\partial}{\partial x} - iea\right)\psi_\pm = 0\,. \tag{67}$$

This completes the demonstration that the hyperspherical formalism is completely equivalent to the usual formulation of quantum electrodynamics in Euclidean x space.

E. Interpretation of Symmetries on the Hypersphere

We briefly discuss in this section the x-space interpretation of the rotational and inversion symmetries on the hypersphere. As we have seen, the photon wave operator is $L^2 - 4$, and this commutes with the ten generators L_{ab} of rotations on the hypersphere. Similarly, the free Dirac wave operator $i\gamma\cdot L + 2$ commutes with the ten operators

$$J_{ab} = L_{ab} + S_{ab}\,, \tag{68}$$

which are the hyperspherical rotation generators when spin is taken into account. We can interpret the hyperspherical rotational symmetry as follows. Six of the rotational generators $L_{\mu\nu}$ (or $J_{\mu\nu}$) leave the 5-axis invariant, and therefore, by Eq. (6b), leave x^2 unchanged. These clearly correspond in x space to the generators of the homogeneous

Lorentz group [which of course, in the Euclidean metric which we use, has become the 4-dimensional rotation group O(4)]. The remaining four generators $L_{5\nu}$, which change x^2, correspond to rather complicated conformal transformations in x space. For example, the 5-dimensional rotation

$$\begin{aligned}&\eta \to \eta':\\ &\eta'_{1,2,3}=\eta_{1,2,3},\\ &\eta'_4=\eta_4\cos\alpha-\eta_5\sin\alpha,\\ &\eta'_5=\eta_4\sin\alpha+\eta_5\cos\alpha\end{aligned} \tag{69a}$$

corresponds in x space to the conformal transformation

$$\begin{aligned}&x\to x':\\ &x'_{1,2,3}=x_{1,2,3}/D,\\ &x'_4=[x_4\cos\alpha-\tfrac{1}{2}(1-x^2)\sin\alpha]/D,\\ &D=1+x^2\sin^2(\tfrac{1}{2}\alpha)+x_4\sin\alpha.\end{aligned} \tag{69b}$$

From this point of view, the manifest covariance of the hyperspherical Feynman rules under rotations generated by $L_{5\nu}$ is a reflection of the conformal invariance of zero fermion-mass electrodynamics. We note, finally, that the ordinary x-space translation $x\to x'=x+a$ *does not* correspond to a linear transformation on η, but rather to the nonlinear transformation

$$\begin{aligned}&\eta\to\eta':\\ &\eta'_\mu=[\eta_\mu+(1+\eta_5)a_\mu]/D',\\ &\eta'_5=[\eta_5-\tfrac{1}{2}a^2(1+\eta_5)-\eta\cdot a]/D',\\ &D'=1+\tfrac{1}{2}a^2(1+\eta_5)+\eta\cdot a,\\ &\eta\cdot a=\eta_\mu a_\mu.\end{aligned} \tag{70}$$

Translation invariance of the x-space formalism guarantees that the 5-dimensional formalism is covariant under the conformal transformations of Eq. (70), even though this is not manifestly evident.

In addition to the continuous-parameter rotation group, there is an important discrete symmetry operation on the hypersphere, the inversion

$$\eta\to-\eta. \tag{71a}$$

According to Eq. (6), this corresponds in x space to the inverse radius transformation

$$x\to -x/x^2. \tag{71b}$$

Because the trace of an odd number of α matrices vanishes, the hyperspherical expression for the closed loop $2n$-point function in Eq. (44) is invariant under simultaneous inversion of all the coordinates $\eta_1,\ldots,\eta_{2n}$. Similarly, the hyperspherical photon propagator is inversion invariant. Hence we conclude that (as long as no divergent vacuum polarization insertions are made) the radiative corrected $2n$-point functions in the hyperspherical formalism are manifestly inversion invariant. This is turn implies simple transformation properties for the corresponding x-space $2n$-point function under simultaneous inverse radius transformation of the coordinates $x_1,\ldots,x_{2n}$. To find the form of the x-space transformation, we follow the notation of Eq. (53), and let $J_a(\eta)$ describe the emission of a photon from the $2n$-point function at coordinate η, with the other $2n-1$ variables suppressed. In this notation, the inversion invariance of the $2n$-point function reads

$$J_a(\eta'=-\eta)=J_a(\eta), \tag{72}$$

where, of course, the suppressed variables are also inverted. Projecting $J_a(\eta)$ back to the x space gives

$$\kappa^{-3}j_\mu(x)=J_\mu(\eta)-x_\mu J_5(\eta), \tag{73}$$

while projecting the inverted current $J_a(\eta')$ gives

$$(\kappa')^{-3}j_\mu(x')=J_\mu(\eta')-x'_\mu J_5(\eta'), \tag{74}$$

with

$$\begin{aligned}&\eta'=-\eta,\\ &x'=1+\eta'_5\\ &\quad=1-\eta_5,\\ &x'_\mu=-x_\mu/x^2.\end{aligned} \tag{75}$$

Using Eqs. (73) and (74), we can convert the equality of Eq. (72) into a relation between $j_\mu(x')$ and $j_\mu(x)$, giving

$$j_\mu(x)=(x^2)^{-3}M_{\mu\nu}(x)j_\nu(x'), \tag{76}$$

where

$$\begin{aligned}&M_{\mu\nu}(x)=\delta_{\mu\nu}-\frac{2x_\mu x_\nu}{x^2},\\ &M_{\mu\nu}(x)M_{\nu\sigma}(x)=\delta_{\mu\sigma}.\end{aligned} \tag{77}$$

Thus, the $2n$-point function in x space is left invariant under the combined operations of (i) simultaneous inverse radius transformation $x_j\to -x_j/x_j^2$, $j=1,\ldots,2n$, and (ii) application of the projection operator

$$\prod_{j=1}^{2n}(x_j^2)^{-3}\prod_{j=1}^{2n}M_{\mu'_j\mu_j}(x_j)$$

to the vector indices. This recipe is just the one discussed by Schreier.[2] In terms of the matrix $M_{\mu\nu}$ we can understand the significance of the gauge terms in the x-space photon propagator of Eq. (56): The gauge terms guarantee that under inverse radius transformations the photon propagator transforms covariantly, i.e.,

$$\Delta_{\mu_1\mu_2}(x_1, x_2) = M_{\mu_1\mu_1'}(x_1)M_{\mu_2\mu_2'}(x_2)\Delta_{\mu_1'\mu_2'}(-x_1/x_1^2, -x_2/x_2^2)\,. \tag{78}$$

The usual Feynman propagator, of course, does not satisfy Eq. (78).

III. CONNECTION WITH THE MANIFESTLY CONFORMAL-COVARIANT FORMALISM

In this section we discuss massless, Euclidean electrodynamics in the manifestly conformal-covariant O(5, 1) language, and develop its connection with the 5-dimensional formalism of the preceding section. In Sec. III A we review the O(5, 1) formalism and in Sec. III B we develop, in a heuristic fashion, the O(5, 1) Feynman rules for electrodynamics. In Sec. III C we show that the O(5, 1) rules are related to the 5-dimensional rules by a simple projective transformation.

A. The O(5,1) Formalism

As has been greatly emphasized recently,[1] a large class of renormalizable field theories containing no dimensional parameters (masses or dimensional coupling constants) are invariant under the 15-parameter conformal group of transformations on space-time. In particular, quantum electrodynamics with zero fermion mass is conformal invariant. We recall that of the 15 conformal-group generators, 10 are the generators of the Poincaré group, 1 generates the dilatations

$$x_\mu \to \lambda x_\mu\,, \tag{79}$$

and the remaining 4 generate the special conformal transformations

$$x_\mu \to \frac{x_\mu + c_\mu x^2}{1 + 2c\cdot x + c^2x^2}\,. \tag{80}$$

Although the usual formulations of massless field theories are manifestly Poincaré-invariant, their invariance under the nonlinear transformations of Eq. (80) is not manifestly evident. However, a very pretty way of achieving manifest conformal invariance was introduced by Dirac,[10] and has been further developed recently. The basic idea is to replace the usual field equations over the Minkowski space-time manifold x_μ by equivalent field equations over a 6-dimensional projective manifold ξ_A. (We adopt the convention that 6-dimensional vector indices are indicated by capital Latin letters $A, B, \ldots$ which take the values $1, \ldots, 6$.) The coordinate x is related to ξ_A by the projective transformation

$$x = \xi_\mu/\xi_+, \quad \xi_+ = \xi_5 + \xi_6\,. \tag{81}$$

When the metric in x space is the Minkowski metric (1, 1, 1, −1), the ξ space is endowed with the metric (1, 1, 1, −1, 1, −1); correspondingly, when the metric in x space is the Euclidean metric (1, 1, 1, 1), the ξ space is endowed with the metric (1, 1, 1, 1, 1, −1). In either case, if ξ is restricted to the light cone

$$\xi^2 = 0\,, \tag{82}$$

then it can be shown that the 15-parameter linear group of pseudorotations on ξ is isomorphic to the conformal group of nonlinear transformations on x. In the Minkowski case, the pseudorotations form the pseudo-orthogonal group O(4, 2), while in the Euclidean case with which we are primarily concerned, they form the pseudo-orthogonal group O(5, 1). So to construct a manifestly conformal invariant formulation of massless, Euclidean electrodynamics, we must write equations which are manifestly covariant under the operations of O(5,1).

Because excellent reviews are available in the literature,[11] we will not actually detail the development of the O(5, 1)-covariant formalism, but rather will simply summarize the results needed for the construction of Feynman rules.

(1) The electromagnetic current is represented by a 6-vector $J_A(\xi)$, homogeneous in ξ of degree −3 and satisfying the kinematic constraint

$$\xi\cdot J(\xi) = 0\,. \tag{83}$$

The equation of electromagnetic current conservation takes the form[9]

$$L_{AB}J^B(\xi) = J_A(\xi)\,, \tag{84a}$$

with

$$L_{AB} = \xi_A\frac{\partial}{\partial\xi^B} - \xi_B\frac{\partial}{\partial\xi^A}\,, \tag{84b}$$

and $J_A(\xi)$ is related to the x-space current $j_\mu(x)$ by the recipe

$$j_\mu(x) = \xi_+^3\{J_\mu(\xi) - x_\mu[J_5(\xi) + J_6(\xi)]\}\,. \tag{85}$$

Note that Eqs. (84) and (85) are both invariant under "gauge" transformations of the form

$$J_A(\xi) \to J_A(\xi) + \xi_A M(\xi)\,, \tag{86}$$

with $M(\xi)$ homogeneous in ξ of degree −4. The invariance of Eq. (85) follows immediately from Eq. (81), while Eq. (84a) is left unchanged because

$$L_{AB}\xi^B M(\xi) = \xi_A\left(5 + \xi\cdot\frac{\partial}{\partial\xi}\right)M(\xi) = \xi_A M(\xi)\,, \tag{87}$$

where in the second equality we have used the homogeneity of $M(\xi)$.

(2) The electromagnetic potential is represented by a 6-vector $A_B(\xi)$, homogeneous in ξ of degree -1 and satisfying the constraint $\xi \cdot A = 0$. The photon wave equation takes the form

$$\Box_6 A_B(\xi) = eJ_B(\xi), \tag{88a}$$

with

$$\Box_6 = \frac{\partial}{\partial \xi_B}\frac{\partial}{\partial \xi^B}. \tag{88b}$$

(3) The electron field is represented by an 8-component spinor $\chi(\xi)$, homogeneous in ξ of degree -2, which obeys the wave equation

$$\left\{ i\gamma_{AB}\left[\xi^A\left(\frac{\partial}{\partial \xi_B} - ieA^B(\xi)\right) - \xi^B\left(\frac{\partial}{\partial \xi_A} - ieA^A(\xi)\right)\right] + 2\right\}\chi = 0. \tag{89}$$

The matrix γ_{AB} is defined by

$$\gamma_{AB} = \tfrac{1}{4}i[\beta_A, \beta_B], \tag{90}$$

where the 8×8 matrices β_A satisfy a Clifford algebra

$$\{\beta_A, \beta_B\} = 2g_{AB} \tag{91}$$

with g_{AB} the metric tensor. An explicit representation of the β's is

$$\beta_\mu = -\gamma_\mu\tau_3, \quad \beta_5 = \tau_1, \quad \beta_6 = -i\tau_2. \tag{92}$$

The electromagnetic current of the electron is given, in terms of the spinor χ, by

$$J_A = 2\xi^B\bar{\chi}\gamma_{BA}\chi, \quad \bar{\chi} = \chi^\dagger\beta_6. \tag{93}$$

These equations completely specify the O(5,1)-covariant formulation of massless electrodynamics, and, via Eq. (85), allow us to project $2n$-point functions in the 6-dimensional language back into $2n$-point functions in x space.

B. O(5,1)-Covariant Feynman Rules

We proceed next to deduce, in a heuristic fashion, Feynman rules for the O(5,1)-covariant calculation of closed-fermion-loop processes. We will not actually directly prove the equivalence of these rules with the usual x-space rules, but rather will show this indirectly in Sec. III C by deducing the 5-dimensional rules of Sec. II from the 6-dimensional rules which we now develop. To begin, we infer from Eq. (93) that the rule for a vertex where a current with polarization index A acts at coordinate ξ is

$$\text{vertex} \propto e\Gamma_A(\xi) = e\xi^B[\beta_B, \beta_A]. \tag{94}$$

Clearly, this rule automatically satisfies the kinematic constraint of Eq. (83). [In Eq. (94) and subsequent equations of the present section, we omit numerical proportionality constants.] Next, we must guess the rule for the electron propagator $S(\xi_1, \xi_2)$. We first note that since $\chi(\xi)$ is homogeneous in ξ of degree -2, S must be homogeneous of degree -2 in ξ_1 and ξ_2 independently. A check on this requirement is provided by the fact that since $J_A(\xi)$ is homogeneous of degree -3, a $2n$-point function must be homogeneous of degree -3 in each of the $2n$ coordinates. Since the vertex $\Gamma_A(\xi)$ is homogeneous of degree $+1$, this requirement will be satisfied by propagator-vertex chains of the form

$$S(\xi_1, \xi_2)\Gamma_{A_2}(\xi_2)S(\xi_2, \xi_3)\Gamma_{A_3}(\xi_3)\cdots \tag{95}$$

only if the propagator is homogeneous of degree -2 in each of its arguments. The homogeneity requirement immediately restricts the choice of propagator to one of two possible forms:

$$S_1(\xi_1, \xi_2) = \frac{\beta\cdot\xi_1\beta\cdot\xi_2}{(\xi_1\cdot\xi_2)^3}, \qquad S_2(\xi_1, \xi_2) = \frac{1}{(\xi_1\cdot\xi_2)^2}. \tag{96}$$

We can rule out S_1 as a possible choice, however, by noting that when S_1 is sandwiched between the two adjacent vertices we get

$$\begin{aligned}\Gamma_{A_1}(\xi_1)S_1(\xi_1, \xi_2)\Gamma_{A_2}(\xi_2) &= \frac{[\beta\cdot\xi_1, \beta_{A_1}]\beta\cdot\xi_1\beta\cdot\xi_2[\beta\cdot\xi_2, \beta_{A_2}]}{(\xi_1\cdot\xi_2)^3} \\ &= -4\frac{\beta\cdot\xi_1(\xi_1)_{A_1}(\xi_2)_{A_2}\beta\cdot\xi_2}{(\xi_1\cdot\xi_2)^3},\end{aligned} \tag{97}$$

where we have used the fact that $(\beta\cdot\xi)^2 = \xi^2 = 0$. But as we have seen above, "gauge" terms of the form $(\xi_1)_{A_1}$ or $(\xi_2)_{A_2}$ project to a null current in x space, so use of S_1 as the propagator would lead to identically vanishing $2n$-point functions in x space. We conclude that the correct choice of electron propagator is S_2, and that the O(5,1)-covariant expression for a closed fermion loop coupling to $2n$ photons is given (up to proportionality constants) by

$$\sum_{\substack{\text{permutations}\\ \text{of } 1,\ldots,2n}} \operatorname{tr}_8\left\{\frac{1}{(\xi_1\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\frac{1}{(\xi_2\cdot\xi_3)^2}\cdots\frac{1}{(\xi_{2n}\cdot\xi_1)^2}[\beta\cdot\xi_1,\beta_{A_1}]\right\}\,. \tag{98}$$

Although we have constructed our rules to satisfy the kinematic constraint of Eq. (83) and the requirements of homogeneity, we must now check whether they are consistent with the equation of current conservation, Eq. (84a). To do this, we first examine the effect of the free Dirac operator on the propagator forms S_1 and S_2. By direct calculation, we find that

$$(i\gamma_{AB}\vec{\mathrm{L}}_1^{AB}+2)S_1(\xi_1,\xi_2)=S_1(\xi_1,\xi_2)(i\gamma_{AB}\overleftarrow{\mathrm{L}}_2^{AB}-2)=0\,, \tag{99a}$$

$$\begin{aligned}&(i\gamma_{AB}\vec{\mathrm{L}}_1^{AB}+2)S_2(\xi_1,\xi_2)=2S_1(\xi_1,\xi_2)\,,\\ &S_2(\xi_1,\xi_2)(i\gamma_{AB}\overleftarrow{\mathrm{L}}_2^{AB}-2)=-2S_1(\xi_1,\xi_2)\,,\end{aligned} \tag{99b}$$

all for $\xi_1\neq\xi_2$ [at $\xi_1=\xi_2$ there are additional δ function contributions, which we omit in writing Eq. (99)]. We see that the correct propagator S_2 *does not* satisfy the Dirac equation, and that adding in an arbitrary multiple of S_1 cannot fix things up. In effect, we see that S_1 is a *null propagator* (because it leads to a vanishing current in x space) and that S_2 is a *pseudopropagator,* which when acted on by the Dirac wave operator gives a multiple of the null propagator, but not zero.

Let us now examine the effect of this peculiar state of affairs on the current conservation properties of Eq. (98). We consider the propagator-vertex chain linking the points ξ_1, ξ, ξ_2 and act with the differential operator $L_{AB}=\xi_A\partial/\partial\xi^B-\xi_B\partial/\partial\xi^A$, giving

$$\begin{aligned}&L_{AB}\cdots[\beta\cdot\xi_1,\beta_{A_1}]\frac{1}{(\xi_1\cdot\xi)^2}[\beta\cdot\xi,\beta^B]\frac{1}{(\xi\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\cdots\\ &\qquad=\cdots[\beta\cdot\xi_1,\beta_{A_1}]\frac{1}{(\xi_1\cdot\xi)^2}[\beta\cdot\xi,\beta_A]\frac{1}{(\xi\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\cdots+R_A\,.\end{aligned} \tag{100}$$

The first term on the right-hand side of Eq. (100) is just the result required by Eq. (84a), while the remainder R_A is given by

$$\begin{aligned}R_A=-2\xi_A\Big\{&\cdots[\beta\cdot\xi_1,\beta_{A_1}]\frac{1}{(\xi_1\cdot\xi)^2}(i\gamma_{AB}\vec{\mathrm{L}}^{AB}+2)\frac{1}{(\xi\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\cdots\\ &+\cdots[\beta\cdot\xi_1,\beta_{A_1}]\frac{1}{(\xi_1\cdot\xi)^2}(i\gamma_{AB}\overleftarrow{\mathrm{L}}^{AB}-2)\frac{1}{(\xi\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\cdots\Big\}\,.\end{aligned} \tag{101}$$

Substituting Eq. (99b) and algebraically rearranging as in Eq. (97), we get[12]

$$R_A=8\xi_A\left\{\cdots[\beta\cdot\xi_1,\beta_{A_1}]\frac{1}{(\xi_1\cdot\xi)^2}\frac{\beta\cdot\xi(\xi_2)_{A_2}\beta\cdot\xi_2}{(\xi\cdot\xi_2)^3}\cdots+\cdots\frac{\beta\cdot\xi_1(\xi_1)_{A_1}\beta\cdot\xi}{(\xi_1\cdot\xi)^3}\frac{1}{(\xi\cdot\xi_2)^2}[\beta\cdot\xi_2,\beta_{A_2}]\cdots\right\}\,. \tag{102}$$

Although Eq. (102) does not vanish, the first term in the curly brackets is a pure "gauge" term with respect to the index A_2, while the second is a pure "gauge" term with respect to the index A_1, and hence both give a vanishing contribution to the $2n$-point function when projected back to x space. So we see that because S_2 is a pseudopropagator, Eq. (98) only satisfies a *pseudocurrent-conservation condition:* When we test current conservation on a given index, Eq. (84a) is *not* satisfied in the 6-dimensional space, but does hold when we project on all of the remaining indices to transform back to x space.

As an explicit illustration of this pseudoconservation property, let us consider the single-loop two-point function, which according to Eq. (98) is given by

$$\frac{\operatorname{tr}_8\{[\beta\cdot\xi_1,\beta_{A_1}][\beta\cdot\xi_2,\beta_{A_2}]\}}{(\xi_1\cdot\xi_2)^4}\propto\frac{\xi_1\cdot\xi_2 g_{A_1A_2}-(\xi_1)_{A_2}(\xi_2)_{A_1}}{(\xi_1\cdot\xi_2)^4}\,. \tag{103}$$

Acting on Eq. (84) with $(L_1)^{A_1'A_1}$ gives

$$\begin{aligned}&(L_1)^{A_1'A_1}\frac{\xi_1\cdot\xi_2 g_{A_1A_2}-(\xi_1)_{A_2}(\xi_2)_{A_1}}{(\xi_1\cdot\xi_2)^4}\\ &\qquad=\frac{\xi_1\cdot\xi_2 g^{A_1'}{}_{A_2}-(\xi_1)_{A_2}(\xi_2)^{A_1'}}{(\xi_1\cdot\xi_2)^4}+R\,,\\ &R=-\frac{4\xi_1\cdot\xi_2(\xi_1)^{A_1'}(\xi_2)_{A_2}}{(\xi_1\cdot\xi_2)^6}\,.\end{aligned} \tag{104}$$

As expected, there is an extra term R which, because it contains the factor $(\xi_2)_{A_2}$, makes no contribution to the two-point function in x space. Interestingly, *there is no way of modifying Eq. (103) to make the extra term R vanish.* To see this we note that the only other second-rank tensor with the correct homogeneity properties and which satisfies the kinematic constraint of Eq. (83) is

$$\frac{(\xi_1)_{A_1}(\xi_2)_{A_2}}{(\xi_1\cdot\xi_2)^4}\,. \tag{105}$$

However, Eq. (87) tells us that this expression satisfies

$$(L_1)^{A_1'A_1}\frac{(\xi_1)_{A_1}(\xi_2)_{A_2}}{(\xi_1\cdot\xi_2)^4}=\frac{(\xi_1)^{A_1'}(\xi_2)_{A_2}}{(\xi_1\cdot\xi_2)^4}\,, \tag{106}$$

so adding a multiple of Eq. (105) to Eq. (103) cannot cancel away R. We conclude that pseudo-current-conservation is an unavoidable feature of the O(5, 1)-covariant formalism.

Next, we turn our attention to the photon propagator $D_{A_1A_2}(\xi_1,\xi_2)$. Because the photon field $A_B(\xi)$ is homogeneous in ξ of degree -1, the photon propagator D must be homogeneous of degree -1 in ξ_1 and ξ_2 independently. In addition, in order to annihilate the extra "gauge" terms which appear when we test current conservation on indices of the closed loop other than A_1, A_2, the photon propagator must be explicitly transverse,

$$(\xi_1)^{A_1}D_{A_1A_2}(\xi_1,\xi_2)=(\xi_2)^{A_2}D_{A_1A_2}(\xi_1,\xi_2)=0\,. \tag{107}$$

The simplest form which satisfies these requirements is

$$D_{A_1A_2}(\xi_1,\xi_2)=\frac{1}{\xi_1\cdot\xi_2}\left[g_{A_1C}-\frac{(\bar\xi_1)_{A_1}(\xi_1)_C}{\bar\xi_1\cdot\xi_1}\right]\left[g^C{}_{A_2}-\frac{(\xi_2)^C(\bar\xi_2)_{A_2}}{\bar\xi_2\cdot\xi_2}\right], \tag{108}$$

where $\bar\xi_1$ and $\bar\xi_2$ are arbitrary points which are held fixed when doing the virtual integrations over ξ_1 and ξ_2. Because of gauge invariance, closed-fermion-loop expressions have no dependence on $\bar\xi_1$ and $\bar\xi_2$ after one sums over all orderings, with respect to other photons which are present, of the emission and absorption of the virtual photon propagated by Eq. (108). The simplest way to verify this statement, and to check the correctness of Eq. (108) to begin with, is to transform Eq. (108) back to x space. We find

$$\xi_{1+}{}^4\xi_{2+}{}^4J^{A_1}(\xi_1)D_{A_1A_2}(\xi_1,\xi_2)J^{A_2}(\xi_2)=-2j_{\mu_1}(x_1)\Delta'_{\mu_1\mu_2}(x_1,x_2)j_{\mu_2}(x_2)\,, \tag{109}$$

with the effective x-space propagator given by

$$\Delta'_{\mu_1\mu_2}(x_1,x_2)=\frac{\delta_{\mu_1\mu_2}}{(x_1-x_2)^2}+2\frac{(\bar x_1-x_1)_{\mu_1}(x_1-x_2)_{\mu_2}}{(\bar x_1-x_1)^2(x_1-x_2)^2}+2\frac{(x_2-x_1)_{\mu_1}(\bar x_2-x_2)_{\mu_2}}{(x_2-x_1)^2(\bar x_2-x_2)^2}-2\frac{(\bar x_1-x_1)_{\mu_1}(\bar x_2-x_2)_{\mu_2}}{(\bar x_1-x_1)^2(\bar x_2-x_2)^2} \tag{110a}$$

$$=\frac{\delta_{\mu_1\mu_2}}{(x_1-x_2)^2}+\frac{\partial}{\partial(x_2)_{\mu_2}}\left[\ln(x_1-x_2)^2\frac{1}{2}\frac{\partial}{\partial(x_1)_{\mu_1}}\ln(\bar x_1-x_1)^2\right]$$

$$+\frac{\partial}{\partial(x_1)_{\mu_1}}\left[\ln(x_1-x_2)^2\frac{1}{2}\frac{\partial}{\partial(x_2)_{\mu_2}}\ln(\bar x_2-x_2)^2\right]-\frac{\partial}{\partial(x_1)_{\mu_1}}\frac{\partial}{\partial(x_2)_{\mu_2}}[\tfrac{1}{2}\ln(\bar x_1-x_1)^2\ln(\bar x_2-x_2)^2]\,. \tag{110b}$$

In Eq. (110a) we give the form of the x-space propagator which emerges directly from the transformation; in this equation $x_1, \bar x_1, x_2, \bar x_2$ denote, respectively, the x-space images of $\xi_1, \bar\xi_1, \xi_2, \bar\xi_2$. In Eq. (110b) we see that Δ' is equivalent, up to gauge terms, to the usual Feynman propagator, and in particular that all the dependence on $\bar x_1, \bar x_2$ is contained in the gauge terms. This verifies that Eq. (108) is a valid expression for the photon propagator, and that $\bar\xi_1$ and $\bar\xi_2$ drop out of gauge invariant quantities, such as closed fermion loops. [The derivation of Eq. (110) from the conformal-covariant expression of Eq. (108) indicates that the effect of the x-space gauge terms is to render Δ' covariant under x-space conformal transformations, provided that the points $\bar x_1, \bar x_2$ are conformally transformed along with x_1 and x_2.[15]]

Finally, calculation of the Jacobian of the transformation of Eq. (81) shows that

$$\int d^4x=\int dS_\xi\,\xi_+{}^{-4}\,, \tag{111}$$

where $\int dS_\xi$ denotes an integration over the hypersphere $\xi_1{}^2+\xi_2{}^2+\xi_3{}^2+\xi_4{}^2+\xi_6{}^2=\xi_6{}^2$, with ξ_6 held

fixed. Comparing with Eq. (109), we see that

$$\int d^4x_1 d^4x_2(-2)j_{\mu_1}(x_1)\Delta'_{\mu_1\mu_2}(x_1,x_2)j_{\mu_2}(x_2) = \int dS_{\xi_1}dS_{\xi_2}J^{A_1}(\xi_1)D_{A_1A_2}(\xi_1,\xi_2)J^{A_2}(\xi_2), \tag{112}$$

indicating that the Feynman rule for virtual integrations is simply

$$\text{virtual integration over } \xi\text{:} \quad \int dS_\xi \,. \tag{113}$$

This completes our specification of the O(5, 1)-covariant Feynman rules for calculating closed-loop quantities.

C. Projection onto the 5-Dimensional Unit Hypersphere

We complete our discussion of the O(5, 1)-covariant formalism by showing that it is related, by a simple projective transformation, to the 5-dimensional Feynman rules of Sec. II. The transformation is generated by exploiting the fact that in an n-virtual photon process, the fixed points $\bar{\xi}$ in each of the n photon propagators can be chosen *independently*, provided that over-all Bose symmetry is maintained. Since closed-fermion-loop amplitudes are independent of all of the propagator fixed points, they will be unchanged if we integrate all of the fixed points over their respective hyperspheres $\bar{\xi}_1^2+\cdots+\bar{\xi}_5^2=\bar{\xi}_6^2$. The effect of this integration is to replace the photon propagator of Eq. (108) by the averaged propagator

$$\bar{D}_{A_1A_2}(\xi_1,\xi_2)=\frac{\int dS_{\bar{\xi}_1}dS_{\bar{\xi}_2}D_{A_1A_2}(\xi_1,\xi_2)}{\int dS_{\bar{\xi}_1}dS_{\bar{\xi}_2}} \,. \tag{114}$$

The integrations in Eq. (114) are readily evaluated, giving

$$\bar{D}_{A_1A_2}(\xi_1,\xi_2)=\frac{1}{\xi_1\cdot\xi_2}\left[g_{A_1C}-\frac{g_{A_16}(\xi_1)_C}{(\xi_1)_6}\right]\left[g^C{}_{A_2}-(\xi_2)^C\frac{g_{A_26}}{(\xi_2)_6}\right]+\text{terms proportional to } (\xi_1)_{A_1} \text{ or } (\xi_2)_{A_2}\,; \tag{115}$$

the terms proportional to $(\xi_1)_{A_1}$ or $(\xi_2)_{A_2}$ are uninteresting because they make a vanishing contribution by virtue of the constraint equation, Eq. (83). The key feature of Eq. (115) is that the quantities in brackets,

$$g_{A_1C}-\frac{g_{A_16}(\xi_1)_C}{(\xi_1)_6}\,,\quad g^C{}_{A_2}-(\xi_2)^C\frac{g_{A_26}}{(\xi_2)_6}\,, \tag{116}$$

both vanish when $C=6$, so the sum in Eq. (115) extends only over $C=1,\ldots,5$. This suggests projecting onto a 5-dimensional space, as follows:

(i) The 5-dimensional coordinate η_a is related to the 6-dimensional coordinate ξ_A by

$$\eta_a=\frac{\xi_a}{\xi_6}\,,\quad a=1,\ldots,5\,. \tag{117a}$$

The light-cone restriction on ξ implies that

$$\eta^2=1\,, \tag{117b}$$

and scalar products in 6-space may be written in 5-space as follows:

$$\xi_1\cdot\xi_2=-\tfrac{1}{2}(\xi_1)_6(\xi_2)_6(\eta_1-\eta_2)^2\,. \tag{117c}$$

(ii) The five-dimensional current $J_a(\eta)$ is related to the 6-dimensional current $J_A(\xi)$ by

$$J_a(\eta)=\xi_6{}^5[J_a(\xi)-\eta_aJ_6(\xi)], \tag{118}$$

which is just the projection generated by the brackets of Eq. (116).

We proceed now to combine Eqs. (115), (117), and (118). Using

$$\int dS_\xi=\int d\Omega_\eta\xi_6{}^4\,, \tag{119}$$

we get

$$\int dS_{\xi_1}dS_{\xi_2}J^{A_1}(\xi_1)\bar{D}_{A_1A_2}(\xi_1,\xi_2)J^{A_2}(\xi_2) = \int d\Omega_{\eta_1}d\Omega_{\eta_2}J_{a_1}(\eta_1)\frac{-2\delta_{a_1a_2}}{(\eta_1-\eta_2)^2}J_{a_2}(\eta_2)\,, \tag{120}$$

which reproduces the 5-dimensional Feynman rule for photon propagation. To study the effect of the projection operation of Eq. (118) on the O(5, 1)-covariant expression for a closed fermion loop in Eq. (98), we consider first the projection of the vertex $\Gamma_A(\xi)$. We find

$$\xi_6{}^3[g_a{}^A - \eta_a g_6{}^A]\Gamma_A(\xi) = \xi_6{}^3[\beta\cdot\xi, \beta_a - \eta_a\beta_6]$$
$$= \xi_6{}^4[\beta\cdot\eta - \beta_6, \beta_a - \eta_a\beta_6]$$
$$= -\xi_6{}^4(\alpha\cdot\eta + 1)\alpha_a(\alpha\cdot\eta - 1)\,, \tag{121}$$

where we have introduced matrices α_a defined by

$$\alpha_a = -\beta_6\beta_a\,. \tag{122}$$

Since the propagator $(\xi_1\cdot\xi_2)^{-2}$ can be rewritten as

$$\frac{1}{(\xi_1\cdot\xi_2)^2} = \frac{4}{(\xi_1)_6{}^2(\xi_2)_6{}^2(\eta_1-\eta_2)^4}\,, \tag{123}$$

we see that the projection of Eq. (118) transforms Eq. (98) into

$$4^{2n}\sum_{\substack{\text{permutations}\\ \text{of } 1,\ldots,2n}} \mathrm{tr}_8\left\{\frac{1}{(\eta_1-\eta_2)^4}(\alpha\cdot\eta_2+1)\alpha_{a_2}(\alpha\cdot\eta_2-1)\frac{1}{(\eta_2-\eta_3)^4}\cdots\frac{1}{(\eta_{2n}-\eta_1)^4}(\alpha\cdot\eta_{2n}+1)\alpha_{a_1}(\alpha\cdot\eta_1-1)\right\}\,, \tag{124}$$

which apart from normalization constants is identical with Eq. (44). So we have verified that the projective transformation generated by using the averaged propagator of Eq. (114) just gives the 5-dimensional Feynman rules for the photon propagator, the electron propagator, and the electron-photon vertex.

To conclude, we show that Eq. (118) and the formal properties of the 6-dimensional current $J_A(\xi)$ imply the corresponding formal properties of the 5-dimensional current $J_a(\eta)$. We begin with the constraint equation, $\xi\cdot J_A(\xi)=0$, which can be rewritten as

$$0 = \xi_a J_a(\xi) - \xi_6 J_6(\xi)$$
$$= \xi_6\eta_a[J_a(\xi) - \eta_a J_6(\xi)]$$
$$= \xi_6{}^{-2}\eta\cdot J(\eta)\,, \tag{125}$$

giving the 5-dimensional constraint equation, Eq. (7). Next we consider the 6-dimensional version of current conservation,

$$L_{AB}J^B(\xi) = J_A(\xi)\,, \tag{126}$$

and use Eq. (87), with $M(\xi)=\xi_6{}^{-1}J_6(\xi)$, to write

$$L_{AB}[J^B(\xi) - \xi^B\xi_6{}^{-1}J_6(\xi)] = J_A(\xi) - \xi_A\xi_6{}^{-1}J_6(\xi)\,. \tag{127}$$

Since the sum on B in Eq. (127) extends only over $B=1,\ldots,5$, and since we are interested only in values of the free index $A=1,\ldots,5$, no derivatives $\partial/\partial\xi_6$ appear. Hence, on multiplying through by $\xi_6{}^S$ and using the fact that

$$\xi_a\frac{\partial}{\partial\xi_b} - \xi_b\frac{\partial}{\partial\xi_a} = \eta_a\frac{\partial}{\partial\eta_b} - \eta_b\frac{\partial}{\partial\eta_a} = L_{ab}\,, \tag{128}$$

Eq. (127) becomes the 5-dimensional current conservation equation

$$L_{ab}J_b(\eta) = J_a(\eta)\,. \tag{129}$$

IV. DISCUSSION

In this section we very briefly discuss possible generalizations and applications of the 5-dimensional formalism of Sec. II. First, we note that although we have worked with a Euclidean x-space metric throughout, it should be straightforward to generalize the 5-dimensional rules to the usual Minkowski case. The hypersphere will then become the hyperbolic domain

$$\eta_1{}^2+\eta_2{}^2+\eta_3{}^2-\eta_4{}^2+\eta_5{}^2=1\,, \tag{130}$$

which is a (4, 1) de Sitter space of unit radius.[14] A further generalization would consist of giving the electron a mass m and, since the distance scale now acquires a meaning, calling the radius of the de Sitter space R, so that Eq. (130) becomes

$$\eta_1{}^2+\eta_2{}^2+\eta_3{}^2-\eta_4{}^2+\eta_5{}^2=R^2\,. \tag{131}$$

As Dirac[5] has shown, an appropriate wave equation describing a massive electron in a de Sitter space of radius R is

$$\left\{i\gamma_{ab}\left[\eta_a\left(\frac{\partial}{\partial\eta_b} - ieA_b(\eta)\right) - \eta_b\left(\frac{\partial}{\partial\eta_a} - ieA_a(\eta)\right)\right] + 2 - imR\right\}\chi = 0\,, \tag{132}$$

which is a simple generalization of Eq. (40). The electron propagator corresponding to Eq. (132) will of course differ from the massless propagator of Table I, but the electron-photon vertex and the photon propagator will be unchanged. The massive 5-dimensional formalism is *not* exactly equivalent to ordinary massive electrodynamics in Minkowski space-time, but as Dirac[5] has shown, in any finite neighborhood of $\eta_5=R$, Eq. (132) reduces to the usual x-space Dirac equation in the limit $R\to\infty$. It is only in the completely massless case that the 5-dimensional and x-space formalisms have the same physical content.

It should be emphasized that while our electron wave equation is identical (in the case $m=0$) to

Dirac's, our treatment of the Maxwell equations is substantially different. Unlike our expressions for the electromagnetic field strengths, which involve $\partial/\partial\eta_a$ only through the angular momentum operator L_{ab}, Dirac's expressions[5] involve $\partial/\partial\eta_a$ by itself. Hence, in order to avoid going off the hypersurface of constant η^2, Dirac finds it necessary to introduce homogeneity constraints on the electromagnetic potential, of the type encountered in the O(5,1)-covariant formalism. In our formulation of the 5-dimensional theory, such constraints are unnecessary, and an examination of the 5-dimensional Feynman rules of Eq. (9) shows, in fact, that they are not homogeneous in the coordinates. The absence of homogeneity requirements permits eigenfunction expansions of the field operators, and should therefore make possible a canonical quantization of the 5-dimensional formalism.[16] The first step in canonical quantization would be to write down an appropriate Lagrangian density; it is readily seen that a variation of

$$\mathcal{L} = -\tfrac{1}{12}(F_{abc})^2 + \bar{\chi}\left\{i\gamma_{ab}\left[\eta_a\left(\frac{\partial}{\partial\eta_b} - ieA_b(\eta)\right) - \eta_b\left(\frac{\partial}{\partial\eta_a} - ieA_a(\eta)\right)\right] + 2 - imR\right\}\chi \tag{133}$$

gives the correct equations of motion. It should then be possible to devise a canonical quantization procedure which reproduces the Feynman rules of Table I from the Lagrangian of Eq. (133). A related question is that of developing the connection between our 5-dimensional formalism and the quantization of electrodynamics by ordering with respect to x^2 which has recently been developed by Del Giudice, Fubini, and Jackiw.[16]

This concludes our discussion of possible avenues for generalization of our results.[17] Let us next briefly consider possible calculational advantages of the 5-dimensional formalism for massless electrodynamics. The key point to notice is that whereas the wave operator in Euclidean x space is $\Box_x^{\,2}$, with a continuum spectrum $-p^2 = -(\text{momentum})^2$, the wave operator on the hypersphere is $L^2 - 4$, with discrete spectrum $-2(n+1)\times(n+2)$. This difference in spectra has two important consequences. First, the fact that the spectrum of $\Box_x^{\,2}$ contains 0 leads to the occurrence of infrared divergences in x-space calculations of propagators and vertex parts. These divergences are known to cancel, however, in closed fermion vacuum polarization loops,[18] and this is reflected in our ability to map vacuum polarization calculations onto the unit hypersphere, where the spectrum of the wave operator does not contain 0. In other words, closed-fermion-loop calculations on the unit hypersphere are manifestly infrared-finite. Second, it is difficult to see how to introduce approximations in massless electrodynamics when calculating in Euclidean x space, since there is no natural scale for selecting one region of p^2 as being more important than another. [We have particularly in mind the calculation of the function $F^{[1]}(\alpha)$ defined in Sec. I, where no natural scale for making approximations is provided by external momenta.] The situation is different on the hypersphere, where unity is a natural scale for measuring the spectrum $-2(n+1)(n+2)$, and where the semiclassical region of large quantum numbers, $n \gg 1$, provides a natural domain for making approximations. The development of techniques for making such semiclassical approximations on the hypersphere is an important problem, which, hopefully, may shed light on the nature of the elusive function $F^{[1]}(\alpha)$.

Added Note

We briefly discuss here two additional topics connected with the 5-dimensional formalism: (a) the photon propagator in the 5-dimensional analog of the Landau gauge, and (b) the hyperspherical harmonic expansion of the 5-dimensional electron propagator.

a. 5-Dimensional Landau Gauge. The x-space photon propagator in the generalized Landau gauge is obtained by adding to the Feynman propagator an appropriate multiple of the gauge term

$$\frac{\partial}{\partial(x_1)_{\mu_1}}\frac{\partial}{\partial(x_2)_{\mu_2}}\ln(x_1 - x_2)^2 = \frac{-2}{(x_1 - x_2)^2}\left[\delta_{\mu_1\mu_2} - 2\,\frac{(x_1 - x_2)_{\mu_1}(x_1 - x_2)_{\mu_2}}{(x_1 - x_2)^2}\right], \tag{A1}$$

adjusted so as to eliminate the logarithmic divergence in the electron wave function renormalization Z_2. Since Eq. (A1) transforms covariantly under inverse radius transformations in x space, we expect that it can be directly transcribed into the 5-dimensional formalism, and indeed a simple calculation shows that

$$j_{\mu_1}(x_1)j_{\mu_2}(x_2)(\kappa_1\kappa_2)^{-4}\frac{1}{(x_1-x_2)^2}\left[\delta_{\mu_1\mu_2}-2\frac{(x_1-x_2)_{\mu_1}(x_1-x_2)_{\mu_2}}{(x_1-x_2)^2}\right]$$

$$=J_{a_1}(\eta_1)J_{a_2}(\eta_2)\frac{1}{(\eta_1-\eta_2)^2}\left[\delta_{a_1a_2}-2\frac{(\eta_1-\eta_2)_{a_1}(\eta_1-\eta_2)_{a_2}}{(\eta_1-\eta_2)^2}\right]. \quad \text{(A2)}$$

Referring now to Eq. (56b) in the text, we note that the gradient terms which guarantee the coordinate-inversion covariance of $\Delta_{\mu_1\mu_2}(x_1x_2)$ behave at worst as $(x_1-x_2)^{-1}$ as $x_1\to x_2$, and hence do not contribute to the logarithmically divergent part of Z_2. In other words, Eq. (56b) is an inversion-covariant form of the Feynman gauge photon propagator. Similarly, corresponding to the usual translation-invariant Landau-gauge photon propagator

$$\frac{\delta_{\mu_1\mu_2}}{(x_1-x_2)^2}+\frac{\lambda}{(x_1-x_2)^2}\left[\delta_{\mu_1\mu_2}-2\frac{(x_1-x_2)_{\mu_1}(x_1-x_2)_{\mu_2}}{(x_1-x_2)^2}\right] \quad \text{(A3)}$$

there is an inversion-covariant Landau gauge photon propagator

$$\Delta_{\mu_1\mu_2}(x_1x_2)+\frac{\lambda}{(x_1-x_2)^2}\left[\delta_{\mu_1\mu_2}-2\frac{(x_1-x_2)_{\mu_1}(x_1-x_2)_{\mu_2}}{(x_1-x_2)^2}\right]. \quad \text{(A4)}$$

Using Eq. (A2) to transcribe Eq. (A4) into the 5-dimensional formalism, we find that the 5-dimensional Landau gauge photon propagator is given by

$$\frac{\delta_{a_1a_2}}{(\eta_1-\eta_2)^2}+\frac{\lambda}{(\eta_1-\eta_2)^2}\left[\delta_{a_1a_2}-\frac{(\eta_1-\eta_2)_{a_1}(\eta_1-\eta_2)_{a_2}}{(\eta_1-\eta_2)^2}\right], \quad \text{(A5)}$$

with λ the *same* constant as appears in the x-space version in Eq. (A3).

b. Electron Propagator Expansion. We consider the 5-dimensional electron propagator, and look for a hyperspherical harmonic expansion of the form

$$\frac{-i}{\pi^2}\frac{\frac{1}{2}(\alpha\cdot\eta_1-1)\frac{1}{2}(\alpha\cdot\eta_2+1)}{(\eta_1-\eta_2)^4}=\tfrac{1}{4}i(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)\sum_{n=0}^{\infty}D_n\sum_m Y_{nm}(\eta_1)Y^*_{nm}(\eta_2)$$

$$=\frac{i}{16\pi^2}(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)\sum_{n=0}^{\infty}(2n+3)D_nC_n^{3/2}(\eta_1\cdot\eta_2). \quad \text{(A6)}$$

Using the orthonormality of the Gegenbauer polynomials $C_n^{3/2}(\mu)$, we find that the expansion coefficients D_n are formally given by the logarithmically divergent expression

$$(2n+3)D_n=-N_n^{-1}\int_{-1}^{1}d\mu(1-\mu^2)\frac{C_n^{3/2}(\mu)}{(1-\mu)^2}, \quad \text{(A7)}$$

with

$$N_n=\int_{-1}^{1}d\mu(1-\mu^2)[C_n^{3/2}(\mu)]^2. \quad \text{(A8)}$$

We proceed by explicitly separating out the logarithmic divergence, writing

$$(2n+3)D_n=A_nI+(2n+3)B_n, \quad \text{(A9)}$$

with

$$A_n=-N_n^{-1}C_n^{3/2}(1)=-\tfrac{1}{4}(2n+3),\quad I=\int_{-1}^{1}d\mu\frac{1+\mu}{1-\mu}, \quad \text{(A10)}$$

and with B_n given by the *convergent* integral

$$(2n+3)B_n=-N_n^{-1}\int_{-1}^{1}d\mu(1-\mu^2)\frac{C_n^{3/2}(\mu)-C_n^{3/2}(1)}{(1-\mu)^2}. \quad \text{(A11)}$$

Substituting Eq. (A9) into Eq. (A6), we see that the contribution of the logarithmically divergent part is

$$\frac{-i}{64\pi^2}(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)\sum_{n=0}^{\infty}(2n+3)C_n^{3/2}(\eta_1\cdot\eta_2)I=\frac{-i}{8}(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_2+1)\delta_S(\eta_1-\eta_2)I\,, \tag{A12}$$

which vanishes since $(\alpha\cdot\eta_1-1)(\alpha\cdot\eta_1+1)=0$. Thus, D_n is effectively given by the integral in Eq. (A11), which can be evaluated by generating function methods. Up to an n-independent piece [which, as we have seen, makes a vanishing contribution to Eq. (A6)], we find

$$D_n=1+\frac{1}{2}+\cdots+\frac{1}{n+1}\,. \tag{A13}$$

ACKNOWLEDGMENTS

We wish to thank H. Abarbanel, D. Gross, and K. Johnson for helpful conversations, and to acknowledge the hospitality of the Aspen Center for Physics, where this work was completed.

---

*Operated by Universities Research Association, Inc., under contract with the U. S. Atomic Energy Commission.

[1]For recent reviews, see G. Mack and A. Salam, Ann. Phys. (N.Y.) 53, 174 (1969); D. J. Gross and J. Wess, Phys. Rev. D 2, 753 (1970); C. G. Callan, S. Coleman, and R. Jackiw, Ann. Phys. (N.Y.) 59, 42 (1970).

[2]E. J. Schreier, Phys. Rev. D 3, 982 (1971); R. J. Crewther, Phys. Rev. Letters 28, 1421 (1972).

[3]This point has been emphasized by M. Baker and K. Johnson (unpublished).

[4]For a review of work on finite electrodynamics, see S. L. Adler, Phys. Rev. D 5, 3021 (1972).

[5]P. A. M. Dirac, Ann. Math. 36, 657 (1935).

[6]As usual, we take $\hbar=c=1$, $e^2/4\pi=\alpha=$ fine-structure constant.

[7]The mapping of Eq. (6a) is closely related to the Fock solution of the hydrogen atom: V. Fock, Z. Physik 98, 145 (1935). For a recent review, see M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 (1966).

[8]The Gegenbauer polynomial and hyperspherical harmonic formulas are obtained from *Higher Transcendental Functions* (Bateman Manuscript Project), edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. II, Sec. 3.15, and Chap. XI; *Tables of Integral Transforms* (Bateman Manuscript Project), edited by A. Erdélyi, *ibid.*, Sec. 16.3; L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*, Translations of Mathematical Monographs, Vol. 6 (American Mathematical Society, Providence, R. I., 1963), Chap. VII.

[9]This form of the current conservation equation was first introduced, in the context of the O(4, 2) conformal-covariant formalism, by D. G. Boulware, L. Brown, and R. D. Peccei, Phys. Rev. D 2, 293 (1970).

[10]P. A. M. Dirac, Ann. Math. 37, 429 (1936).

[11]See G. Mack and A. Salam, Ref. 1; D. G. Boulware *et al.*, Ref. 9.

[12]We have again omitted δ-function contributions.

[13]Propagators of the form Eq. (91a), with $\bar{x}_1=\bar{x}_2$, have been studied by M. Baker and K. Johnson (unpublished) and by R. A. Abdellatif, Ph.D. thesis, University of Washington, 1970 (unpublished).

[14]For a good review see F. Gürsey, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964). For an exhaustive bibliography on field theories in de Sitter space, see S. A. Fulling, Ph.D. thesis, Princeton University, 1972 (unpublished).

[15]For discussions of quantization of scalar field theories in de Sitter space, see S. A. Fulling, Ref. 14, and references quoted therein, especially M. Gutzwiller, Helv. Phys. Acta 29, 313 (1956).

[16]Del Giudice, S. Fubini, and R. Jackiw (unpublished).

[17]For completeness, we note that the techniques which we have developed in this paper for the case of massless electrodynamics will be applicable to other conformally invariant field theories as well.

[18]T. Kinoshita, J. Math. Phys. 3, 650 (1962); T. D. Lee and M. Nauenberg, Phys. Rev. 133B, 1549 (1964).

Massless, Euclidean Quantum Electrodynamics on the 5-Dimensional Unit Hypersphere. Stephen L. Adler [Phys. Rev. D 6, 3445 (1972)]. 1. Page 3447, Table I. The normalizing factors for the 5-dimensional and Euclidean electron propagators should read $(-1/\pi^2)$ and $(-1/2\pi^2)$, respectively, instead of $(-i/\pi^2)$ and $(-i/2\pi^2)$ as given. (Corresponding changes should be made elsewhere in the paper.) 2. Page 3449, first column, third line following Eq. (44): Eq. (8a) should read Eq. (8).

PHYSICAL REVIEW D VOLUME 10, NUMBER 8 15 OCTOBER 1974

Massless electrodynamics in the one-photon-mode approximation

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 6 May 1974)

We discuss single-fermion-loop vacuum-polarization processes in massless quantum electrodynamics in the one-photon-mode approximation, in which the fermion self-interacts (to all orders in perturbation theory) by the exchange of virtual photons in a single virtual-photon eigenmode. The isolation of one photon mode is made possible by using the O(5)-covariant formulation of massless QED introduced in two earlier papers, in which the photon wave operator has a discrete, rather than a continuous, spectrum. The amplitude integral formalism introduced previously expresses the one-mode radiative-corrected vacuum polarization in terms of the uncorrected vacuum amplitude in the presence of a one-mode external field. By exploiting the residual $SO(3)\times O(2)$ symmetry of the one-mode external-field problem, which permits separation of variables, we reduce the external-field problem to a set of two coupled ordinary first-order differential equations. We show that when the two independent solutions to these equations are suitably standardized, their Wronskian gives (up to a constant factor) the external-field-problem Fredholm determinant. We study the distribution of zeros and asymptotic behavior of the Fredholm determinant, relate these properties to the coupling-constant analyticity of the one-mode vacuum polarization, and conclude by giving a brief list of unresolved questions.

I. INTRODUCTION

We begin in this paper the analysis of a simple, nonperturbative approximation to single-fermion-loop vacuum-polarization processes in massless quantum electrodynamics. In our approximation, the virtual fermion in the vacuum-polarization loop self-interacts to all orders of perturbation theory only by the exchange of virtual photons in a single virtual-photon eigenmode. The isolation of one photon mode is made possible by using the O(5)-covariant formulation of massless QED introduced in two earlier papers,[1,2] in which the photon wave operator has a discrete, rather than a continuous, spectrum. Specifically, our approximation is obtained by replacing the full effective photon propagator

$$D^{(0)}_{ab}(\eta_1,\eta_2)_{\rm eff}=\sum_{n,m}\frac{Y^{(1)}_{nma}(\eta_1)Y^{(1)}_{nmb}(\eta_2)}{(n+1)(n+2)} \tag{1.1}$$

by the simple, factorizable form

$$\hat{D}_{ab}(\eta_1,\eta_2)=\tfrac{1}{6}Y^{(1)}_{1Ma}(\eta_1)Y^{(1)}_{1Mb}(\eta_2), \tag{1.2}$$

which results when the sum in Eq. (1.1) is truncated to contain only *one* of the 10 modes in the smallest ($n=1$) photon representation of O(5). Specifically, the one mode which we retain has the form

$$Y^{(1)}_{1Ma}(\eta)=\left(\frac{15}{16\pi^2}\right)^{1/2}(v_{1a}\eta\cdot v_2-v_{2a}\eta\cdot v_1), \tag{1.3}$$

where v_1 and v_2 are arbitrary, orthogonal five-dimensional unit vectors,

$$v_1^{\,2}=v_2^{\,2}=1,\quad v_1\cdot v_2=0, \tag{1.4}$$

and where η is the five-dimensional coordinate.

As was shown in Ref. 2, the radiative-corrected single-fermion-loop vacuum functional in the one-mode approximation (denoted by $W_1[A']$) is given by the amplitude integral formula

$$W_1[a'Y^{(1)}_{1M}/e]=\int_{-\infty}^{\infty}da\left(\frac{6}{2\pi e^2}\right)^{1/2}\exp\left(\frac{-3a^2}{e^2}\right)\times W^{(0)}[(a+a')Y^{(1)}_{1M}], \tag{1.5}$$

where $W^{(0)}[A]$ is the single-fermion-loop vacuum functional in the presence of an external electromagnetic potential A, with *no* internal-virtual-photon radiative corrections (and with the dependence on the electric charge e eliminated by a rescaling of the electromagnetic potential). Formally, $W^{(0)}[A]$ is given by the expression

$$W^{(0)}[A]=\tfrac{1}{2}\operatorname{Tr}\ln h_T,\qquad h_T=2-L\cdot S-i\alpha\cdot\eta\alpha\cdot A, \tag{1.6}$$

with the anticommuting matrices α and the O(5) angular momentum and spin L and S defined as in Ref. 2. If we introduce the eigenvalues μ of h_T (which, as we shall see, occur in quadruples μ, μ, $-\mu$, $-\mu$) and define the external-field-problem Fredholm determinant

$$\Delta[A]=\Big(\prod_{\text{all eigenvalues}}\mu\Big)^{1/4}, \tag{1.7}$$

then $W^{(0)}$ can be written as

$$W^{(0)}[A]=2\ln\Delta[A]. \tag{1.8}$$

As is evident from Eqs. (1.5)–(1.8) and as was developed in detail in Ref. 2, the analyticity properties of W_1 as a function of coupling e^2 are deter-

mined by the asymptotic behavior of $W^{(0)}[aY^{(1)}_{1M}]$ for large external-field amplitude a, or, what is essentially equivalent, by the distribution of zeros of the Fredholm determinant Δ in the complex a plane.

Let us now spell out more specifically the connection between the e^2 analyticity of W_1 and the a dependence of $W^{(0)}$. In order to make Eq. (1.5) unambiguous, we must specify the integration contour to be used in evaluating the a integral. In Ref. 2 we argued that this contour should be taken to be along the real a axis, or possibly (and very conjecturally) along the imaginary a axis. Equation (1.5) with real integration contour will be well defined if Δ has no zeros (and hence $W^{(0)}$ no singularities) for a real. If $W^{(0)}$ is asymptotically weaker than an increasing Gaussian in a as a becomes infinite along the real axis, then Eq. (1.5) defines an analytic function of e^2 in the right-hand e^2 half plane. If, moreover, Δ has no singularities in the wedge-shaped sectors $|\mathrm{Re}a|>|\mathrm{Im}a|$ and the vacuum amplitude $W^{(0)}$ is asymptotically weaker than a Gaussian in these sectors, the integration contour can be freely deformed within these sectors from its original position along the real axis, implying that W_1 is an analytic function of e^2 in the entire e^2 plane, apart from a branch cut along the negative real axis from $e^2=0$ to $e^2=-\infty$. Thus, for real integration contour the questions at stake are:

(i) Is Δ zero-free for a real?

(ii) Is $W^{(0)}$ asymptotically weaker than a Gaussian as $a\to\pm\infty$ along the real axis?

(iii) Is Δ zero-free in the sectors $|\mathrm{Re}a|>|\mathrm{Im}a|$?

(iv) Is $W^{(0)}$ asymptotically weaker than a Gaussian as $|a|\to\infty$ within the sectors?

In the following sections we present analytic arguments which answer questions (i), (ii), and (iv) in the affirmative, and we present numerical results (but no proofs) which also suggest an affirmative answer for question (iii). Next let us consider the speculative possibility of an imaginary integration contour. Such a contour is allowed only if *two* conditions are satisfied: Δ must have no zeros for purely imaginary a, and $W^{(0)}$ must vanish as a decreasing Gaussian (or faster) as $a\to\infty$ along the imaginary axis. As shown in Ref. 2, if $W^{(0)}$ oscillates along the imaginary axis with a decreasing Gaussian envelope, then the imaginary contour yields a strong-coupling electrodynamics in which W_1 exists for large enough e^2 and can develop an infinite-order zero as e^2 approaches a positive e_0^2 from above. Thus, the questions at issue for a possible imaginary integration contour are:

(v) Is Δ zero-free for a imaginary?

(vi) What is the asymptotic behavior of $W^{(0)}$ as $|a|\to\infty$ along the imaginary axis?

The analytic arguments which follow answer question (v) affirmatively. With respect to question (vi) we can only give limited numerical results, these show no signs of decreasing asymptotic behavior, but, because the asymptotic region may not have been reached, do not conclusively resolve question (vi).

The material which follows is organized so that a knowledge of the O(5) formalism is needed only to read Sec. II, in which we consider the wave equation determining the eigenvalues μ of h_T,

$$[2-L\cdot S-ia\alpha\cdot\eta\alpha\cdot Y^{(1)}_{1M}(\eta)]\psi=\mu\psi\,, \tag{1.9}$$

and show that separation of variables with respect to the SO(3)×O(2) subgroup of O(5) reduces Eq. (1.9) to a pair of coupled ordinary first-order differential equations within each separable subspace. In the remaining sections, which can be read independently of Sec. II, we study the properties of this differential-equation system. In Sec. III we recapitulate the results of Sec. II and argue directly from the differential equations that Δ has no zeros for a in strips containing the real and imaginary axes. In Sec. IV we construct the Green's function of the one-dimensional system, and use it to establish a connection between the Wronskian of the two independent solutions of the differential equations (suitably standardized) and the Fredholm determinant Δ. In Sec. V we use this connection, combined with WKB estimates, to determine the order of growth of Δ for large $|a|$. In Sec. VI we construct series solutions for the two independent solutions of the differential equation, and use them to study $\Delta(a)$ numerically. Finally, in Sec. VII we briefly summarize the many remaining unresolved questions. In Appendix A we explicitly calculate the Green's function in the free case, and in Appendix B we give the details of the WKB calculation used in Sec. V.

II. REDUCTION OF THE ONE-MODE PROBLEM

In this section we carry out the separation of variables which reduces the partial differential equation (1.9) to a pair of coupled ordinary first-order differential equations. In Sec. II A we determine the conserved quantum numbers of Eq. (1.9), and show that the eigenvalue problem diagonalizes with respect to an SO(3)×O(2) subgroup of O(5). In Sec. II B we introduce a representation of the O(5) generators which facilitates reduction of the eigenvalue problem with respect to the conserved subgroup. The reduction itself is carried out in Sec. II C. In Sec. II D, we perform a check by solving the free ($a=0$) case and verifying the eigenvalue degeneracies found in Ref. 2. We also

work out the boundary conditions appropriate to the separated equations in both the free and the interacting cases. Finally, in Sec. II E we make a transformation which simplifies the equations in the interacting case, and construct the external field problem Fredholm determinant introduced in Sec. I.

A. Conserved quantum numbers

To analyze the conserved quantum numbers of Eq. (1.9) we choose axes in the five-dimensional space so that the 1 and 2 axes lie, respectively, along v_1 and v_2. The Hamiltonian in Eq. (1.9) then takes the form

$$h_T = h_T^{(0)} + V,$$
$$h_T^{(0)} = 2 - L \cdot S, \quad V = i\lambda\alpha\cdot\eta(\alpha_1\eta_2 - \alpha_2\eta_1), \tag{2.1}$$
$$\lambda = -a(15/16\pi^2)^{1/2}.$$

Introducing the O(5) generators

$$J_{ab} = L_{ab} + S_{ab}$$
$$= \eta_a \frac{\partial}{\partial\eta_b} - \eta_b \frac{\partial}{\partial\eta_a} + \tfrac{1}{4}[\alpha_a, \alpha_b] \tag{2.2}$$

we obviously have

$$[J_{ab}, h_T^{(0)}] = 0 \tag{2.3}$$

since the free Hamiltonian $h_T^{(0)}$ is rotationally invariant. Furthermore, since

$$[L_{ab}, \eta_c] = \eta_a\delta_{bc} - \eta_b\delta_{ac},$$
$$[S_{ab}, \alpha_c] = \alpha_a\delta_{bc} - \alpha_b\delta_{ac}, \tag{2.4}$$

we find, as expected, that $\alpha\cdot\eta$ is also rotationally invariant,

$$[J_{ab}, \alpha\cdot\eta] = 0. \tag{2.5}$$

Hence the generators J_{ab} which commute with h_T will be just the ones which commute with the factor $\alpha_1\eta_2 - \alpha_2\eta_1$ in the potential term. From Eq. (2.4) we find trivially that

$$[J_{34}, \alpha_1\eta_2 - \alpha_2\eta_1] = [J_{35}, \alpha_1\eta_2 - \alpha_2\eta_1]$$
$$= [J_{45}, \alpha_1\eta_2 - \alpha_2\eta_1]$$
$$= 0, \tag{2.6}$$

indicating that h_T is invariant under the SO(3) subgroup generated by J_{34}, J_{35}, and J_{45}. In addition, we have

$$[J_{12}, \alpha_1\eta_2 - \alpha_2\eta_1] = -\alpha_2\eta_2 - \alpha_1\eta_1 + \alpha_1\eta_1 + \alpha_2\eta_2$$
$$= 0, \tag{2.7}$$

so that h_T is also invariant under the O(2) subgroup generated by J_{12}. The other generators J_{ab} do not commute with h_T. In addition to the SO(3) $\times$O(2) invariance group which we have just found, there are also two discrete invariances of h_T. Defining a coordinate inversion generator P,

$$P\eta P^{-1} = -\eta, \quad P^2 = 1, \tag{2.8}$$

we see immediately that

$$[P, h_T] = 0. \tag{2.9}$$

Finally, letting α_6 be the α matrix which anticommutes with $\alpha_1, \ldots, \alpha_5$, we have

$$[\alpha_6, h_T] = 0. \tag{2.10}$$

This last invariance permits us to split the eight-component spinor eigenvalue problem of Eq. (1.9) into two identical decoupled four-component problems. Diagonalizing the four-component spinor with respect to the conserved quantum numbers, we write

$$\psi = \psi_{jm\xi\epsilon},$$
$$(J_{34}{}^2 + J_{35}{}^2 + J_{45}{}^2)\psi_{jm\xi\epsilon} = -j(j+1)\psi_{jm\xi\epsilon},$$
$$J_{45}\psi_{jm\xi\epsilon} = im\psi_{jm\xi\epsilon}, \tag{2.11}$$
$$J_{12}\psi_{jm\xi\epsilon} = i\xi\psi_{jm\xi\epsilon},$$
$$P\psi_{jm\xi\epsilon} = \epsilon\psi_{jm\xi\epsilon}.$$

As we will see in detail below, the separation constants take the values

$$j = \tfrac{1}{2}, \tfrac{3}{2}, \ldots,$$
$$m = -j, -j+1, \ldots, j, \tag{2.12}$$
$$\xi = \pm\tfrac{1}{2}, \pm\tfrac{3}{2}, \ldots,$$
$$\epsilon = \pm 1.$$

Our task in the succeeding sections will be to find the form taken by the eigenvalue problem of Eq. (1.9) when restricted to the subspace of Eq. (2.11).

B. Explicit representation of the O(5) generators

We introduce now an explicit representation of the O(5) generators which facilitates the reduction of the eigenvalue problem with respect to the SO(3)$\times$O(2) subgroup. We begin with the spin operator $S_{ab} = \frac{1}{4}[\alpha_a, \alpha_b]$. Letting $\sigma_{1,2,3}$, $\tau_{1,2,3}$, and $\rho_{1,2,3}$ be three commuting sets of 2×2 Pauli spin matrices, we represent the 8×8 matrices $\alpha_1, \ldots, \alpha_6$ in the form

$$\alpha_1 = \sigma_1\tau_2, \quad \alpha_2 = \sigma_2\tau_2, \quad \alpha_3 = \rho_3\sigma_3\tau_2,$$
$$\alpha_4 = \rho_1\sigma_3\tau_2, \quad \alpha_5 = \rho_2\sigma_3\tau_2, \quad \alpha_6 = \tau_3, \tag{2.13}$$

so that the spin matrices become

$$S_{12}=\tfrac{1}{2}i\sigma_3,\quad S_{14}=-\tfrac{1}{2}i\rho_1\sigma_2,$$
$$S_{24}=\tfrac{1}{2}i\rho_1\sigma_1,\quad S_{53}=\tfrac{1}{2}i\rho_1,$$
$$S_{15}=-\tfrac{1}{2}i\rho_2\sigma_2,\quad S_{25}=\tfrac{1}{2}i\rho_2\sigma_1, \tag{2.14}$$
$$S_{34}=\tfrac{1}{2}i\rho_2,\quad S_{13}=-\tfrac{1}{2}i\rho_3\sigma_2,$$
$$S_{23}=\tfrac{1}{2}i\rho_3\sigma_1,\quad S_{45}=\tfrac{1}{2}i\rho_3.$$

Since the Hamiltonian h_T is even in the α matrices $\alpha_1,\ldots,\alpha_5$, it is a unit matrix in the space of the τ Pauli matrices. As noted above, this immediately reduces Eq. (1.9) to two identical decoupled four-component eigenvalue problems.

To represent the orbital angular momentum L_{ab}, we parametrize the coordinates $\eta_1,\ldots,\eta_5$ in the form

$$\eta_1=\sin\theta_1\cos\phi_1,\quad \eta_2=\sin\theta_1\sin\phi_1,$$
$$\eta_3=\cos\theta_1\cos\theta_2,\quad \eta_4=\cos\theta_1\sin\theta_2\cos\phi_2,$$
$$\eta_5=\cos\theta_1\sin\theta_2\sin\phi_2, \tag{2.15}$$
$$0\le\theta_1\le\tfrac{1}{2}\pi,\quad 0\le\phi_1\le 2\pi,$$
$$0\le\theta_2\le\pi,\quad 0\le\phi_2\le 2\pi$$

corresponding to an O(2) (angle ϕ_1) and an SO(3) (angles θ_2,ϕ_2) combined with mixing angle θ_1. In terms of these angular parameters, the coordinate inversion operation is

$$\eta\to-\eta \Leftrightarrow \begin{cases}\theta_1\to\theta_1, & \phi_1\to\phi_1+\pi,\\ \theta_2\to\pi-\theta_2, & \phi_2\to\phi_2+\pi.\end{cases} \tag{2.15'}$$

The hyperspherical surface element becomes

$$d\Omega_\eta=\det\begin{bmatrix}\eta_1 & \frac{\partial\eta_1}{\partial\theta_1} & \frac{\partial\eta_1}{\partial\phi_1} & \frac{\partial\eta_1}{\partial\theta_2} & \frac{\partial\eta_1}{\partial\phi_2}\\ \eta_2 & & & & \\ \cdot & & & & \cdot\\ \cdot & & & & \cdot\\ \cdot & & & & \cdot\\ \eta_5 & \frac{\partial\eta_5}{\partial\theta_1} & \cdots & & \frac{\partial\eta_5}{\partial\phi_2}\end{bmatrix} d\theta_1 d\phi_1 d\theta_2 d\phi_2$$
$$=\cos^2\theta_1\sin\theta_1 d\theta_1(d\phi_1)(\sin\theta_2 d\theta_2 d\phi_2), \tag{2.16}$$

which, not surprisingly, has a mixing-angle factor, an O(2) factor $(d\phi_1)$, and an SO(3) factor $(\sin\theta_2 d\theta_2 d\phi_2)$. By dint of considerable algebra one can express the orbital angular momenta in terms of derivatives with respect to the angles of Eq. (2.15). To write the results in a compact form, we introduce auxiliary operators M_j, N_j, P_j, $j=1,\ldots,3$, as follows:

$$M_1=-\sin\phi_1\frac{\partial}{\partial\theta_1}-\cot\theta_1\cos\phi_1\frac{\partial}{\partial\phi_1},$$
$$M_2=\cos\phi_1\frac{\partial}{\partial\theta_1}-\cot\theta_1\sin\phi_1\frac{\partial}{\partial\phi_1},$$
$$M_3=\frac{\partial}{\partial\phi_1},$$
$$N_1=-\sin\phi_2\frac{\partial}{\partial\theta_2}-\cot\theta_2\cos\phi_2\frac{\partial}{\partial\phi_2},$$
$$N_2=\cos\phi_2\frac{\partial}{\partial\theta_2}-\cot\theta_2\sin\phi_2\frac{\partial}{\partial\phi_2}, \tag{2.17}$$
$$N_3=\frac{\partial}{\partial\phi_2},$$
$$P_1=\cos\theta_2\cos\phi_2\frac{\partial}{\partial\theta_2}-\csc\theta_2\sin\phi_2\frac{\partial}{\partial\phi_2},$$
$$P_2=\cos\theta_2\sin\phi_2\frac{\partial}{\partial\theta_2}+\csc\theta_2\cos\phi_2\frac{\partial}{\partial\phi_2},$$
$$P_3=-\sin\theta_2\frac{\partial}{\partial\theta_2}.$$

These satisfy the commutation relations and identities

$$\left.\begin{array}{l}[M_i,M_j]=-M_k,\quad [M_i,N_j]=[M_i,P_j]=0,\\ [N_i,N_j]=-N_k,\\ [P_i,P_j]=N_k,\quad [N_i,P_j]=-P_k,\end{array}\right\} i,j,k \text{ cyclic}$$
$$\vec{N}^2=\vec{P}^2=\frac{1}{\sin\theta_2}\frac{\partial}{\partial\theta_2}\left(\sin\theta_2\frac{\partial}{\partial\theta_2}\right)+\frac{1}{\sin^2\theta_2}\frac{\partial^2}{\partial\phi_2{}^2}. \tag{2.18}$$

In terms of the auxiliary operators, the orbital angular momentum operators take the form

$$L_{12}=M_3,\quad L_{53}=N_1,$$
$$L_{34}=N_2,\quad L_{45}=N_3,$$
$$L_{14}=-\sin\theta_2\cos\phi_2M_2+\tan\theta_1\cos\phi_1P_1,$$
$$L_{15}=-\sin\theta_2\sin\phi_2M_2+\tan\theta_1\cos\phi_1P_2, \tag{2.19}$$
$$L_{13}=-\cos\theta_2M_2+\tan\theta_1\cos\phi_1P_3,$$
$$L_{24}=\sin\theta_2\cos\phi_2M_1+\tan\theta_1\sin\phi_1P_1,$$
$$L_{25}=\sin\theta_2\sin\phi_2M_1+\tan\theta_1\sin\phi_1P_2,$$
$$L_{23}=\cos\theta_2M_1+\tan\theta_1\sin\phi_1P_3,$$

and by using Eqs. (2.17) and (2.18) it is straightforward to verify that the expressions in Eq. (2.19) satisfy the O(5) commutation relations

$$[L_{ab},L_{cd}]=\delta_{ac}L_{db}-\delta_{ad}L_{cb}+\delta_{bc}L_{ad}-\delta_{bd}L_{ac}. \tag{2.20}$$

Using Eqs. (2.14) and (2.19), it is a simple matter to express the Hamiltonian h_T and the conserved generators J_{12}, J_{53}, ... in terms of the angular parameters. We find

$$\begin{aligned} h_T^{(0)} &= 2 - i[M_3\sigma_3 + N_1\rho_1 + N_2\rho_2 + N_3\rho_3 + (M_1\sigma_1 + M_2\sigma_2)(\rho_1 \sin\theta_2 \cos\phi_2 + \rho_2 \sin\theta_2 \sin\phi_2 + \rho_3\cos\theta_2) \\ &\quad + i\tan\theta_1 \sigma_3(\sigma_1\cos\phi_1 + \sigma_2\sin\phi_1)(P_1\rho_1 + P_2\rho_2 + P_3\rho_3)] , \\ V &= \lambda \sin\theta_1[\sigma_3 \sin\theta_1 - \cos\theta_1(\sigma_1\cos\phi_1 + \sigma_2 \sin\phi_1)(\rho_1 \sin\theta_2\cos\phi_2 + \rho_2\sin\theta_2\sin\phi_2 + \rho_3\cos\theta_2)] , \end{aligned} \tag{2.21}$$

and

$$\begin{aligned} U_3 &\equiv -iJ_{12} = -iM_3 + \tfrac{1}{2}\sigma_3 , \\ T_1 &\equiv -iJ_{53} = -iN_1 + \tfrac{1}{2}\rho_1 , \\ T_2 &\equiv -iJ_{34} = -iN_2 + \tfrac{1}{2}\rho_2 , \\ T_3 &\equiv -iJ_{45} = -iN_3 + \tfrac{1}{2}\rho_3 . \end{aligned} \tag{2.22}$$

C. Reduction of the eigenvalue problem

The first step in the reduction of the eigenvalue problem with respect to the SO(3)×O(2) subgroup is to find the eigenvalues and eigenfunctions of the conserved generators in Eq. (2.22). This is, of course, just a standard angular momentum problem. For the O(2) subgroup we find two eigenfunctions with opposite inversion parity for each eigenvalue ξ of U_3,

$$\begin{aligned} U_3 u_\pm &= \xi u_\pm , \\ P u_\pm &= (-1)^{\xi \mp 1/2} u_\pm , \\ u_+ &= e^{i(\xi - 1/2)\phi_1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}_\sigma , \\ u_- &= e^{i(\xi + 1/2)\phi_1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}_\sigma . \end{aligned} \tag{2.23}$$

The subscript σ on the spinors indicates that they are acted on by the Pauli matrices σ_j. Because the orbital angular momentum $-iM_3$ must have integral eigenvalues, the eigenvalues of U_3 must be half-integral; hence the allowed values of ξ are

$$\xi = \pm\tfrac{1}{2},\ \pm\tfrac{3}{2},\ \ldots . \tag{2.24}$$

For the SO(3) subgroup we again find two eigenfunctions with opposite inversion parity for each pair of $\vec{T}$ eigenvalues j, m,

$$\begin{aligned} \vec{T}^2 v_\pm &= j(j+1)v_\pm , \quad T_3 v_\pm = m v_\pm , \\ P v_\pm &= (-1)^{j+2m \mp 1/2} v_\pm , \\ v_+ &= \begin{bmatrix} (j - m + 1)P_{j+1/2}^{m-1/2}(\cos\theta_2)e^{i(m-1/2)\phi_2} \\ P_{j+1/2}^{m+1/2}(\cos\theta_2)e^{i(m+1/2)\phi_2} \end{bmatrix}_\rho , \\ v_- &= \begin{bmatrix} (j + m)P_{j-1/2}^{m-1/2}(\cos\theta_2)e^{i(m-1/2)\phi_2} \\ -P_{j-1/2}^{m+1/2}(\cos\theta_2)e^{i(m+1/2)\phi_2} \end{bmatrix}_\rho , \end{aligned} \tag{2.25}$$

with $P_L^M(z)$ the usual associated Legendre polynomial. The allowed values of j, m are the usual ones for spin-$\frac{1}{2}$ coupled to an orbital angular momentum,

$$\begin{aligned} j &= \tfrac{1}{2}, \tfrac{3}{2}, \ldots , \\ m &= -j,\ -j+1, \ldots, j , \end{aligned} \tag{2.26}$$

and the subscript ρ indicates that the spinors are acted on by the Pauli matrices ρ_j. In terms of the O(2) and SO(3) eigenfunctions which we have just found, the general decomposition of $\psi_{jm\xi\pm}$ is

$$\begin{aligned} \psi_{jm\xi\epsilon} &= A_+(\theta_1)v_+u_+ + C_+(\theta_1)v_-u_- , \\ \psi_{jm\xi-\epsilon} &= A_-(\theta_1)v_+u_- + C_-(\theta_1)v_-u_+ , \\ \epsilon &= (-1)^{\xi+j+2m-1} . \end{aligned} \tag{2.27}$$

The next step is to substitute Eq. (2.27) into Eq. (1.9), using the expression of Eq. (2.21) for h_T. To find the action of the various terms of h_T on $u_\pm$ and $v_\pm$ we use the following identities, which may be verified by straightforward calculation:

$$\begin{aligned} &\sigma_3 u_\pm = \pm u_\pm , \\ &(\sigma_1\cos\phi_1 + \sigma_2\sin\phi_1)u_\pm = u_\mp , \\ &(2 - iM_3\sigma_3)u_\pm = (\tfrac{3}{2} \pm \xi)u_\pm , \\ &-i(M_1\sigma_1 + M_2\sigma_2)u_\pm = \pm u_\mp\left[\frac{d}{d\theta_1} + (\tfrac{1}{2} \mp \xi)\cot\theta_1\right] ; \\ &(\rho_1\sin\theta_2\cos\phi_2 + \rho_2\sin\theta_2\sin\phi_2 + \rho_3\cos\theta_2)v_\pm = v_\mp , \\ &-i\vec{N}\cdot\vec{\rho}v_+ = -(j+\tfrac{3}{2})v_+ , \quad -i\vec{N}\cdot\vec{\rho}v_- = (j-\tfrac{1}{2})v_- , \\ &\vec{P}\cdot\vec{\rho}v_+ = (j+\tfrac{3}{2})v_- , \quad \vec{P}\cdot\vec{\rho}v_- = -(j-\tfrac{1}{2})v_+ . \end{aligned} \tag{2.28}$$

Hence we get

$$\begin{aligned}
h_T\psi_{jm\xi\epsilon} &= (\tfrac{3}{2}+\xi)A_+v_+u_+ + (\tfrac{3}{2}-\xi)C_+v_-u_- - (j+\tfrac{3}{2})A_+v_+u_+ + (j-\tfrac{1}{2})C_+v_-u_- \\
&\quad + \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}-\xi)\cot\theta_1\right]A_+v_-u_- - \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}+\xi)\cot\theta_1\right]C_+v_+u_+ \\
&\quad - \tan\theta_1(j+\tfrac{3}{2})A_+v_-u_- - \tan\theta_1(j-\tfrac{1}{2})C_+v_+u_+ + \lambda\sin^2\theta_1 A_+v_+u_+ - \lambda\sin^2\theta_1 C_+v_-u_- \\
&\quad - \lambda\sin\theta_1\cos\theta_1 A_+v_-u_- - \lambda\sin\theta_1\cos\theta_1 C_+v_+u_+ \\
&= \mu\psi_{jm\xi\epsilon} \\
&= \mu A_+v_+u_+ + \mu C_+v_-u_-\,, \\
h_T\psi_{jm\xi-\epsilon} &= (\tfrac{3}{2}-\xi)A_-v_+u_- + (\tfrac{3}{2}+\xi)C_-v_-u_+ - (j+\tfrac{3}{2})A_-v_+u_- + (j-\tfrac{1}{2})C_-v_-u_+ \\
&\quad - \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}+\xi)\cot\theta_1\right]A_-v_-u_+ + \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}-\xi)\cot\theta_1\right]C_-v_+u_- \\
&\quad + \tan\theta_1(j+\tfrac{3}{2})A_-v_-u_+ + \tan\theta_1(j-\tfrac{1}{2})C_-v_+u_- - \lambda\sin^2\theta_1 A_-v_+u_- + \lambda\sin^2\theta_1 C_-v_-u_+ \\
&\quad - \lambda\sin\theta_1\cos\theta_1 A_-v_-u_+ - \lambda\sin\theta_1\cos\theta_1 C_-v_+u_- \\
&= \mu\psi_{jm\xi-\epsilon} \\
&= \mu A_-v_+u_- + \mu C_-v_-u_+\,.
\end{aligned} \tag{2.29}$$

Equating coefficients of like terms then gives us the following two sets of coupled first-order differential equations for $A_\pm(\theta_1)$ and $C_\pm(\theta_1)$:

$$\begin{aligned}
&(\xi-j)A_+ - \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}+\xi)\cot\theta_1 + (j-\tfrac{1}{2})\tan\theta_1\right]C_+ + \lambda\sin\theta_1(A_+\sin\theta_1 - C_+\cos\theta_1) = \mu A_+\,, \\
&(j+1-\xi)C_+ + \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}-\xi)\cot\theta_1 - (j+\tfrac{3}{2})\tan\theta_1\right]A_+ - \lambda\sin\theta_1(C_+\sin\theta_1 + A_+\cos\theta_1) = \mu C_+\,;
\end{aligned} \tag{2.30a}$$

$$\begin{aligned}
&-(\xi+j)A_- + \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}-\xi)\cot\theta_1 + (j-\tfrac{1}{2})\tan\theta_1\right]C_- - \lambda\sin\theta_1(\sin\theta_1 A_- + \cos\theta_1 C_-) = \mu A_-\,, \\
&(j+1+\xi)C_- - \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}+\xi)\cot\theta_1 - (j+\tfrac{3}{2})\tan\theta_1\right]A_- + \lambda\sin\theta_1(\sin\theta_1 C_- - \cos\theta_1 A_-) = \mu C_-\,.
\end{aligned} \tag{2.30b}$$

These two sets of equations can be further reduced to just one set of coupled differential equations by exploiting the fact that

$$\alpha\cdot\eta h_T = -h_T\alpha\cdot\eta\,. \tag{2.31}$$

Since $\alpha\cdot\eta$ has odd inversion parity, Eq. (2.31) tells us that if $\psi_{jm\xi-\epsilon}$ is an eigenfunction of h_T with eigenvalue μ, then $\alpha\cdot\eta\psi_{jm\xi-\epsilon}$ is an eigenfunction of h_T with eigenvalue $-\mu$, quantum numbers j, m, ξ unaltered, but (reversed) inversion parity $+\epsilon$. Specifically, writing[3]

$$\begin{aligned}
\alpha\cdot\eta\psi_{jm\xi-\epsilon} &= [(\sigma_1\cos\phi_1 + \sigma_2\sin\phi_1)\sin\theta_1 + \sigma_3(\rho_1\sin\theta_2\cos\varphi_2 + \rho_2\sin\theta_2\sin\varphi_2 + \rho_3\cos\theta_2)\cos\theta_1] \\
&\quad \times[A_-(\theta_1)v_+u_- + C_-(\theta_1)v_-u_+] \\
&= \hat{A}_+(\theta_1)v_+u_+ + \hat{C}_+(\theta_1)v_-u_-\,,
\end{aligned} \tag{2.32}$$

we find from the relations of Eq. (2.28) that

$$\begin{aligned}
\hat{A}_+ &= A_-\sin\theta_1 + C_-\cos\theta_1\,, \\
\hat{C}_+ &= -A_-\cos\theta_1 + C_-\sin\theta_1\,.
\end{aligned} \tag{2.33}$$

From the differential equations [Eqs. (2.30b)] satisfied by A_- and C_-, we find that $\hat{A}_+$ and $\hat{C}_+$ satisfy the coupled differential equations

$$\begin{aligned}
&(\xi-j)\hat{A}_+ - \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}+\xi)\cot\theta_1 + (j-\tfrac{1}{2})\tan\theta_1\right]\hat{C}_+ \\
&\qquad + \lambda\sin\theta_1(\hat{A}_+\sin\theta_1 - \hat{C}_+\cos\theta_1) = -\mu\hat{A}_+\,, \\
&(j+1-\xi)\hat{C}_+ + \left[\frac{d}{d\theta_1} + (\tfrac{1}{2}-\xi)\cot\theta_1 - (j+\tfrac{3}{2})\tan\theta_1\right]\hat{A}_+ \\
&\qquad - \lambda\sin\theta_1(\hat{C}_+\sin\theta_1 + \hat{A}_+\cos\theta_1) = -\mu\hat{C}_+\,.
\end{aligned} \tag{2.34}$$

As expected, these are identical to Eqs. (2.30a), apart from the reversal in sign of the eigenvalue. Thus, we need only study the one set of equations in Eq. (2.30a).

To find the measure with respect to which two eigensolutions of Eqs. (2.30a) with different eigenvalues μ, μ' are orthogonal, we start from the hyperspherical orthonormality condition

$$\int d\Omega_\eta \psi^\dagger_{jm\xi\epsilon}\psi'_{jm\xi\epsilon}=0\,,\quad \mu\neq\mu'\,. \tag{2.35}$$

Using the expression for $d\Omega_\eta$ in Eq. (2.16), and the fact that

$$\begin{aligned} u_-^\dagger u_- &= u_+^\dagger(\sigma_1\cos\phi_1+\sigma_2\sin\phi_1)^2u_+\\ &=u_+^\dagger u_+\,,\\ v_-^\dagger v_- &= v_+^\dagger(\rho_1\sin\theta_2\cos\phi_2+\rho_2\sin\theta_2\sin\phi_2\\ &\quad +\rho_3\cos\theta_2)^2v_+\\ &=v_+^\dagger v_+\,,\end{aligned} \tag{2.36}$$

Eq. (2.35) reduces to

$$\int_0^{\pi/2}\cos^2\theta_1\sin\theta_1\,d\theta_1(A_+^*A_+'+C_+^*C_+')=0\,,\quad \mu\neq\mu' \tag{2.37}$$

which identifies the measure for Eqs. (2.30a).

Now that the eigenvalue problem has been reduced to a single set of coupled first-order differential equations, the subscripts used in the above analysis are no longer needed. To expedite the subsequent discussion, let us change notation as follows:

$$\begin{aligned}&\theta_1\rightarrow\theta\,,\\ &A_+(\theta_1)\rightarrow a(\theta)\,,\\ &C_+(\theta_1)\rightarrow c(\theta)\,.\end{aligned} \tag{2.38}$$

The differential equations which we must study thus are

$$\begin{aligned}&(\xi-j)a-\left[\frac{d}{d\theta}+(\tfrac{1}{2}+\xi)\cot\theta+(j-\tfrac{1}{2})\tan\theta\right]c\\ &\qquad +\lambda\sin\theta(a\sin\theta-c\cos\theta)=\mu a\,,\\ &(j+1-\xi)c+\left[\frac{d}{d\theta}+(\tfrac{1}{2}-\xi)\cot\theta-(j+\tfrac{3}{2})\tan\theta\right]a\\ &\qquad -\lambda\sin\theta(c\sin\theta+a\cos\theta)=\mu c\,,\end{aligned} \tag{2.39}$$

with the measure for orthogonality

$$\int_0^{\pi/2}\cos^2\theta\sin\theta\,d\theta(a^*a'+c^*c')=0\,,\qquad \mu\neq\mu'\,. \tag{2.40}$$

D. Solution of the free ($\lambda=0$) case and check on eigenvalue counting

Let us now check the reduction leading to Eq. (2.39) by solving the differential equations in the case of vanishing interaction and comparing the energy spectrum with the free-particle spectrum calculated in Ref. 2. When $\lambda=0$, the differential equations simplify to

$$\begin{aligned}&(\xi-j)a-\left[\frac{d}{d\theta}+(\tfrac{1}{2}+\xi)\cot\theta+(j-\tfrac{1}{2})\tan\theta\right]c=\mu a\,,\\ &(j+1-\xi)c+\left[\frac{d}{d\theta}+(\tfrac{1}{2}-\xi)\cot\theta-(j+\tfrac{3}{2})\tan\theta\right]a=\mu c\,.\end{aligned} \tag{2.41}$$

Changing the independent variable to $u=\cos^2\theta$ and eliminating either c or a, we find that a satisfies a second-order differential equation of standard Riemann type,[4]

$$\frac{d^2a}{du^2}+\left(\frac{3}{2}\,\frac{1}{u}+\frac{1}{u-1}\right)\frac{da}{du}+\left[-\frac{(j+\frac{3}{2})(j+\frac{1}{2})}{4}\,\frac{1}{u^2}-\frac{(\xi-\frac{1}{2})^2}{4}\,\frac{1}{(u-1)^2}+\frac{(j+\frac{3}{2})(j+\frac{1}{2})+(\xi-\frac{1}{2})^2+2+\mu(1-\mu)}{4}\,\frac{1}{u(u-1)}\right]a=0\,, \tag{2.42}$$

and c satisfies a similar equation obtained from Eq. (2.42) by the replacements $j\rightarrow j-1$, $\xi\rightarrow\xi+1$. The characteristic exponents of Eq. (2.42) at the regular singular points at $u=0$ and $u=1$ are given in Table I. Equation (2.42) can be solved in terms of Jacobi polynomials, giving the following four series of eigenfunctions and eigenvalues.

$\xi\geq\frac{1}{2}$:

$$a=f(\cos\theta)^{j+1/2}(\sin\theta)^{\xi-1/2}\times P_n^{(j+1,\xi-1/2)}(1-2\cos^2\theta),$$
$$c=(\cos\theta)^{j-1/2}(\sin\theta)^{\xi+1/2}\times P_n^{(j,\xi+1/2)}(1-2\cos^2\theta).$$

Series 1. (2.43)

$\mu=2n+2+j+\xi,\quad n=0,1,2,\ldots$

$f=-(n+\xi+\frac{1}{2})/(n+j+1)$.

Series 2.

$\mu=-(2n+1+j+\xi),\quad n=0,1,2,\ldots$

$f=1$.

$\xi\leq-\frac{1}{2}$:

$$a=f(\cos\theta)^{j+1/2}(\sin\theta)^{1/2-\xi}\times P_n^{(j+1,1/2-\xi)}(1-2\cos^2\theta),$$
$$c=(\cos\theta)^{j-1/2}(\sin\theta)^{-1/2-\xi}\times P_{n+1}^{(j,-1/2-\xi)}(1-2\cos^2\theta).$$

Series 3. (2.44)

$\mu=2n+3+j-\xi,\quad n=-1,0,1,\ldots$

$f=-1$.

Series 4.

$\mu=-(2n+2+j-\xi),\quad n=0,1,2,\ldots$

$f=(n+j-\xi+\frac{3}{2})/(n+1)$.

These solutions can be verified by direct substitution into Eq. (2.41), using the following four identities satisfied by the Jacobi polynomial $P_n^{(\alpha,\beta)}(x)$[5]:

$$(1-x)\frac{d}{dx}P_n^{(\alpha,\beta)}(x)=\alpha P_n^{(\alpha,\beta)}(x)-(n+\alpha)P_n^{(\alpha-1,\beta+1)}(x),$$
$$(1+x)\frac{d}{dx}P_n^{(\alpha,\beta)}(x)=(n+\beta)P_n^{(\alpha+1,\beta-1)}(x)-\beta P_n^{(\alpha,\beta)}(x),$$
$$\beta(1-x)P_n^{(\alpha,\beta)}(x)-(n+\alpha)(1+x)P_n^{(\alpha-1,\beta+1)}(x)=-2(n+1)P_{n+1}^{(\alpha-1,\beta-1)}(x),\tag{2.45}$$
$$2\frac{d}{dx}P_{n+1}^{(\alpha,\beta)}(x)=(n+\alpha+\beta+2)P_n^{(\alpha+1,\beta+1)}(x).$$

Let us now count the total degeneracy with which the eigenvalue $\mu=k+2$ occurs. Remembering that we have reduced our problem to a four-component spinor, the expected degeneracy of the eigenvalue $\mu=k+2$ is

$$\deg(k+2)=\dim(k+\tfrac{1}{2},\tfrac{1}{2})=\tfrac{2}{3}(k+1)(k+2)(k+3),\quad k=0,1,2,\ldots.\tag{2.46}$$

For each eigenfunction with eigenvalue μ and inversion parity ϵ obtained from Eqs. (2.43) and (2.44), there is another eigenfunction with eigenvalue $-\mu$ and opposite inversion parity obtained by inverting the transformation of Eq. (2.33) to give

$$A_-=a\sin\theta-c\cos\theta,\quad C_-=a\cos\theta+c\sin\theta.\tag{2.47}$$

Hence the positive eigenvalues of h_T are

$$\left.\begin{array}{l}2n+2+j+|\xi|\\2n+1+j+|\xi|\end{array}\right\}\text{twice each}$$
$$n=0,1,2,\ldots,\quad j=\tfrac{1}{2},\tfrac{3}{2},\ldots,\tag{2.48}$$
$$m=-j,\ldots,j,\quad |\xi|=\tfrac{1}{2},\tfrac{3}{2},\ldots,$$

and the degeneracy of the eigenvalue $k+2$ is

$$\deg(k+2)=2\sum_{\substack{n,j,|\xi|\\2n+j+|\xi|=k}}(2j+1)+2\sum_{\substack{n,j,|\xi|\\2n+j+|\xi|=k+1}}(2j+1).\tag{2.49}$$

The right-hand side of Eq. (2.49) is obviously a cubic polynomial in k, which by direct enumeration of the two sums, takes the values 4, 16, 40, 80 for $k=0,1,2,3$, respectively. Hence it is equal to $\frac{2}{3}(k+1)(k+2)(k+3)$, and the eigenvalue-counting checks. In group-theoretic language, what we have done is to exhibit the decomposition of the $(k+\frac{1}{2},\frac{1}{2})$ representation of O(5) in terms of states labeled by the quantum numbers of the SO(3)×O(2) subgroup.

From Eqs. (2.43) and (2.44), we see that in the free case the two-component wave function

$$\psi=\begin{pmatrix}a\\c\end{pmatrix}\tag{2.50a}$$

satisfies the finiteness boundary condition

TABLE I. Characteristic exponents of the differential equations for a and c at $u=\cos^2\theta=0,1$. [See the discussion following Eq. (2.50).]

Singular point: $u=0,\ \theta=\frac{1}{2}\pi$			Singular point: $u=1,\ \theta=0$		
$a\sim u^{\sigma_a}=(\cos\theta)^{2\sigma_a}$			$a\sim(1-u)^{\chi_a}=(\sin\theta)^{2\chi_a}$		
$c\sim u^{\sigma_c}=(\cos\theta)^{2\sigma_c}$			$c\sim(1-u)^{\chi_c}=(\sin\theta)^{2\chi_c}$		
	Characteristic exponents			Characteristic exponents	
	Solution 1	Solution 2		Solution 1	Solution 2
σ_a	$\frac{1}{2}(j+\frac{1}{2})$	$-\frac{1}{2}(j+\frac{3}{2})$	χ_a	$\frac{1}{2}(\xi-\frac{1}{2})$	$-\frac{1}{2}(\xi-\frac{1}{2})$
σ_c	$\frac{1}{2}(j-\frac{1}{2})$	$-\frac{1}{2}(j+\frac{1}{2})$	χ_c	$\frac{1}{2}(\xi+\frac{1}{2})$	$-\frac{1}{2}(\xi+\frac{1}{2})$

$$\psi \sim \text{finite at } \theta=0,\ \theta=\tfrac{1}{2}\pi, \tag{2.50b}$$

and an examination of the characteristic exponents in Table I shows that Eq. (2.50b) is equivalent to the square-integrability boundary condition

$$\int_0^{\pi/2} \cos^2\theta \sin\theta\, d\theta(|a|^2+|c|^2)<\infty\,. \tag{2.50c}$$

Since the interaction term in Eq. (2.39) is nonsingular at $\theta=0$, $\theta=\frac{1}{2}\pi$, the characteristic exponents of the differential equation system at $\theta=0$, $\theta=\frac{1}{2}\pi$ are the same in the interacting case as in the noninteracting case. Hence the boundary condition in Eq. (2.50), which we inferred from the free solution, is appropriate to the interacting case as well.

E. Reduction of the interacting case and construction of the Fredholm determinant

It is convenient, for the work which follows, to reduce the coupled differential equations of Eq. (2.39) to a somewhat simpler form. We work with the two-component spinor notation of Eq. (2.50a), and write Eq. (2.39) in the matrix form

$$H\psi=\mu\psi\,. \tag{2.51}$$

Introducing Pauli matrices τ_1, τ_2, τ_3 which act on the spinor ψ, it is easy to see that H may be written as

$$\begin{aligned} H &= \tfrac{1}{2}-(j-\xi-\lambda\sin^2\theta+\tfrac{1}{2})\tau_3 \\ &\quad -[\xi\cot\theta+(j+\tfrac{1}{2})\tan\theta+\lambda\sin\theta\cos\theta]\tau_1 \\ &\quad -i\left(\frac{d}{d\theta}+\tfrac{1}{2}\cot\theta-\tan\theta\right)\tau_2\,. \end{aligned} \tag{2.52}$$

We now make a similarity transformation on Eqs. (2.51) and (2.52), writing

$$\begin{aligned} \psi &= S\psi_R\,, \\ H &= SH_RS^{-1}\,, \\ S &= (\cos\tfrac{1}{2}\theta-i\tau_2\sin\tfrac{1}{2}\theta)[(\sin\theta)^{1/2}\cos\theta]^{-1}\,. \end{aligned} \tag{2.53}$$

The transformed eigenvalue problem is

$$\begin{aligned} H_R\psi_R &= \mu\psi_R\,, \\ H_R &= -[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\tau_1 \\ &\quad -(j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3-i\frac{d}{d\theta}\tau_2\,. \end{aligned} \tag{2.54}$$

The measure for orthogonality is now

$$\int_0^{\pi/2} d\theta\,\psi_R^\dagger\psi_R'=0,\quad \mu\neq\mu' \tag{2.55}$$

and the boundary condition is

$$\psi_R\sim(\sin\theta)^{1/2}\cos\theta\times\text{finite at } \theta=0,\ \theta=\tfrac{1}{2}\pi\,, \tag{2.56a}$$

or equivalently

$$\int_0^{\pi/2} d\theta|\psi_R|^2<\infty\,. \tag{2.56b}$$

To construct the external-field-problem Fredholm determinant, we display the parameter dependence of the eigenvalue μ by writing

$$\mu=\mu_{\xi j}(\lambda)\,, \tag{2.57}$$

so that the Fredholm determinant within the separable subspace takes the form

$$\begin{aligned} \Delta_{\xi j}(\lambda) &= \prod_{\substack{\text{all eigenvalues}\\ \text{in subspace}}} \mu_{\xi j}(\lambda) \\ &= \det[H_R]\,. \end{aligned} \tag{2.58}$$

Remembering that for each eigenvalue $\mu_{\xi j}(\lambda)$ there is an eigenvalue $-\mu_{\xi j}(\lambda)$ coming from eigenfunctions with opposite inversion parity [see the discussion following Eq. (2.31)], and that there is an additional duplication of eigenvalues when we reconstruct back to eight-component spinors, we find that

$$\prod_{\text{all eigenvalues}} \mu = \prod_{j=1/2,3/2,\ldots}\ \prod_{\xi=\pm1/2,\pm3/2,\ldots} \times \prod_{\substack{\text{all eigenvalues}\\ \text{in subspace}}} [\mu_{\xi j}(\lambda)^4]^{2j+1}\,. \tag{2.59}$$

Thus, comparing with Eq. (1.7), we see that the full external-field-problem Fredholm determinant is given by

$$\Delta[A]=\prod_{j=1/2,3/2,\ldots}\ \prod_{\xi=\pm1/2,\pm3/2,\ldots} \Delta_{\xi j}(\lambda)^{2j+1}\,. \tag{2.60}$$

One further transformation of this formula proves to be useful. From Eq. (2.54), we see that if

$$H_R\psi=\mu_{\xi j}(\lambda)\psi\,, \tag{2.61}$$

then

$$H_R\Big|_{\substack{\xi\to-\xi\\ \lambda\to-\lambda}}\tau_1\psi=-\mu_{\xi j}(\lambda)\tau_1\psi\,. \tag{2.62}$$

Since $\tau_1{}^2=1$, we conclude that the sets of numbers $\{-\mu_{\xi j}(\lambda)\}$, $\{\mu_{-\xi j}(-\lambda)\}$ are identical. Hence

$$\prod_{\substack{\text{all eigenvalues}\\ \text{in subspace}}} \mu_{-\xi j}(\lambda)=\prod_{\substack{\text{all eigenvalues}\\ \text{in subspace}}} (-)\mu_{\xi j}(-\lambda)\,, \tag{2.63}$$

permitting us to eliminate the negative $-\xi$ factors in Eq. (2.60). Dividing out $\Delta[0]$ to eliminate an irrelevant (and infinite) constant factor, we get finally

$$\frac{\Delta[A]}{\Delta[0]}=\prod_{j=1/2,3/2,\ldots}\ \prod_{\xi=1/2,3/2,\ldots}\left[\frac{\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)}{\Delta_{\xi j}(0)^2}\right]^{2j+1}\,. \tag{2.64}$$

Equation (2.64) is still a formal expression, in that renormalizations have not yet been made. In Sec. V below we discuss the modification of Eq. (2.64) which is made necessary by renormalization subtractions, and which guarantees convergence of the infinite product.

III. ZERO-FREE STRIPS

For the benefit of the reader who has skipped Sec. II, we briefly recapitulate the principal results derived there. In terms of the effective external-field amplitude

$$\lambda = -a\left(\frac{15}{16\pi^2}\right)^{1/2}, \tag{3.1}$$

we found that the external-field problem could be reduced to the two-component eigenvalue problem ($\tau_{1,2,3}$ = Pauli matrices)

$$\begin{aligned} H\psi &= \mu_{\xi j}(\lambda)\psi\,, \\ H &= -[\xi(\sin\theta)^{-1} + \lambda\sin\theta]\tau_1 \\ &\quad - (j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3 - i\frac{d}{d\theta}\tau_2\,, \end{aligned} \tag{3.2}$$

with the measure for orthogonality

$$\int_0^{\pi/2} d\theta\,\psi^\dagger\psi' = 0, \quad \mu \neq \mu' \tag{3.3}$$

and the boundary condition

$$\psi \sim (\sin\theta)^{1/2}\cos\theta \times \text{finite at } \theta = 0, \tfrac{1}{2}\pi\,. \tag{3.4}$$

Defining the Fredholm determinant corresponding to Eq. (3.2) by

$$\Delta_{\xi j}(\lambda) = \prod_{\text{all eigenvalues}} \mu_{\xi j}(\lambda)\,, \tag{3.5}$$

we found that the full external-field-problem Fredholm determinant introduced in Eq. (1.7) is given (up to renormalization subtractions) by

$$\frac{\Delta[A]}{\Delta[0]} = \prod_{j=1/2,3/2,\ldots}\;\prod_{\xi=1/2,3/2,\ldots}\left[\frac{\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)}{\Delta_{\xi j}(0)^2}\right]^{2j+1}. \tag{3.6}$$

The reminder of this paper is devoted to a study of the mathematical properties of Eqs. (3.2)–(3.6).

We begin by showing that $\Delta_{\xi j}(\lambda)$ cannot vanish in strips in the λ plane containing the real and imaginary axes. From Eq. (3.5) we see that zeros of $\Delta_{\xi j}(\lambda)$ occur at values of λ where Eq. (3.2) has a vanishing eigenvalue, that is, where

$$H\psi = 0 \tag{3.7}$$

for nonvanishing, normalizable ψ. To get our first restriction on the locations of zeros, we multiply Eq. (3.7) by $\psi^\dagger\tau_1$ and integrate, giving

$$\begin{aligned} &-\int_0^{\pi/2}[\xi(\sin\theta)^{-1} + \lambda\sin\theta]\psi^\dagger\psi d\theta + iR_1 = 0\,, \\ &R_1 = (j+\tfrac{1}{2})\int_0^{\pi/2}(\cos\theta)^{-1}\psi^\dagger\tau_2\psi d\theta \\ &\qquad + \int_0^{\pi/2}\psi^\dagger\tau_3\left(-i\frac{d}{d\theta}\right)\psi d\theta\,. \end{aligned} \tag{3.8}$$

Using the boundary condition of Eq. (3.4) to integrate by parts, we readily see that R_1 is pure real. Hence taking the real part of Eq. (3.8) gives the relation

$$\frac{-\text{Re}\lambda}{\xi} = \frac{\int_0^{\pi/2}(\sin\theta)^{-1}\psi^\dagger\psi d\theta}{\int_0^{\pi/2}\sin\theta\,\psi^\dagger\psi d\theta} \geq 1\,. \tag{3.9}$$

We learn from this relation that $\Delta[A]$ has no zeros for λ in the strip $|\text{Re}\lambda| \leq \frac{1}{2}$, and in particular no zeros on the imaginary axis. To get a second restriction on the locations of zeros, we multiply Eq. (3.7) by $\psi^\dagger\tau_3$ and integrate, giving

$$\begin{aligned} &-i\int_0^{\pi/2}\lambda\sin\theta\,\psi^\dagger\tau_2\psi d\theta \\ &\qquad - (j+\tfrac{1}{2})\int_0^{\pi/2}(\cos\theta)^{-1}\psi^\dagger\psi d\theta + iR_2 = 0\,, \end{aligned} \tag{3.10}$$

$$R_2 = -\xi\int_0^{\pi/2}(\sin\theta)^{-1}\psi^\dagger\tau_2\psi d\theta - \int_0^{\pi/2}\psi^\dagger\tau_1\left(-i\frac{d}{d\theta}\right)\psi d\theta\,.$$

Again, the boundary condition of Eq. (3.4) implies that R_2 is real, so taking the real part of Eq. (3.10) gives the second relation

$$\frac{\text{Im}\lambda}{j+\frac{1}{2}} = \frac{\int_0^{\pi/2}(\cos\theta)^{-1}\psi^\dagger\psi d\theta}{\int_0^{\pi/2}\sin\theta\,\psi^\dagger\tau_2\psi d\theta}\,. \tag{3.11}$$

Since τ_2 has eigenvalues ± 1, we have the inequality $|\psi^\dagger\tau_2\psi| \leq \psi^\dagger\psi$, and so Eq. (3.11) implies the inequality

$$\frac{|\text{Im}\lambda|}{j+\frac{1}{2}} \geq \frac{\int_0^{\pi/2}(\cos\theta)^{-1}\psi^\dagger\psi d\theta}{\int_0^{\pi/2}\sin\theta\,\psi^\dagger\psi d\theta} \geq 1\,. \tag{3.12}$$

Thus $\Delta[A]$ can have no zeros for λ in the strip $|\text{Im}\lambda| \leq 1$, and in particular no zeros on the real axis. Combining the restrictions of Eqs. (3.9) and (3.12), we get the regions in the λ plane where $\Delta_{\xi j}(\lambda)$ is allowed to have zeros, as illustrated in Fig. 1. Note that the absolute value sign in Eq. (3.12) cannot be removed. In fact, since the Hamiltonian H is Hermitian for real λ, $\Delta_{\xi j}(\lambda)$ is a real analytic function of λ and satisfies the reflectien principle

$$\Delta_{\xi j}(\lambda)^* = \Delta_{\xi j}(\lambda^*)\,. \tag{3.13}$$

Hence for each zero λ of $\Delta_{\xi j}(\lambda)$, there is a corresponding zero at the complex-conjugate point λ^*.

IV. WRONSKIAN FORMULA FOR THE FREDHOLM DETERMINANT

We proceed next to derive a connection between the Fredholm determinant in each separable subspace and the Wronskian of two suitably standardized independent solutions of Eq. (3.7). In Sec. IV A we construct the Green's function for H, introduce the standard solutions, and discuss their analyticity and rate of growth in λ. In Sec. IV B we prove the connection between the Wronskian and the Fredholm determinant.

A. Green's function and standard solutions

Let $H=H(\theta)$ be the Hamiltonian of Eq. (3.2), and let $S=H^{-1}$ be the Green's function satisfying

$$H(\theta_1)S(\theta_1,\theta_2)=\delta(\theta_1-\theta_2)1\,,\qquad 0\le\theta_1,\ \theta_2\le\tfrac{1}{2}\pi\,, \tag{4.1}$$

with 1 the 2×2 unit matrix. To construct an explicit expression for S, we introduce the solutions ψ_1,ψ_2 of Eq. (3.7) which are regular at $\theta=0$, $\theta=\frac{1}{2}\pi$, respectively:

$$\begin{aligned} &H\psi_1=H\psi_2=0\,,\\ &\psi_1=\begin{pmatrix}a_1\\ c_1\end{pmatrix}\sim(\sin\theta)^{1/2}\times\text{finite at }\theta=0\,,\\ &\psi_2=\begin{pmatrix}a_2\\ c_2\end{pmatrix}\sim\cos\theta\times\text{finite at }\theta=\tfrac{1}{2}\pi\,. \end{aligned}\tag{4.2}$$

We also need the Wronskian of the two solutions, defined by (the superscript T denotes transpose)

$$w(\lambda)=\psi_2^T i\tau_2\psi_1=a_2(\theta)c_1(\theta)-a_1(\theta)c_2(\theta)\,. \tag{4.3}$$

Since

$$\begin{aligned}\frac{dw}{d\theta}&=\psi_2^T\left(i\tau_2\frac{d}{d\theta}\psi_1\right)-\left(i\tau_2\frac{d}{d\theta}\psi_2\right)^T\psi_1\\ &=-\psi_2^T\{[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\tau_1+(j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3\}\psi_1+\psi_2^T\{[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\tau_1+(j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3\}^T\psi_1=0\,,\end{aligned}\tag{4.4}$$

the Wronskian is θ-independent. Applying the method of variation of parameters,[6] we then find the following expression for S:

$$S(\theta_1,\theta_2)=w^{-1}\times\begin{cases}\psi_1(\theta_1)\psi_2^T(\theta_2)\,, & \theta_1<\theta_2\\ \psi_2(\theta_1)\psi_1^T(\theta_2)\,, & \theta_1>\theta_2\,.\end{cases}\tag{4.5}$$

To verify Eq. (4.5), we note that

$$\begin{aligned}&H(\theta_1)S(\theta_1,\theta_2)=0,\quad\theta_1<\theta_2,\ \theta_1>\theta_2\,;\\ &\int_{\theta_2-\epsilon}^{\theta_2+\epsilon}d\theta_1 H(\theta_1)S(\theta_1,\theta_2)\underset{\epsilon\to 0}{\longrightarrow}-i\tau_2 w^{-1}\\ &\qquad\times[\psi_2(\theta_2)\psi_1^T(\theta_2)-\psi_1(\theta_2)\psi_2^T(\theta_2)]\\ &\qquad=\begin{pmatrix}0&-1\\1&0\end{pmatrix}w^{-1}\begin{pmatrix}0&w\\-w&0\end{pmatrix}\\ &\qquad=1\,,\end{aligned}\tag{4.6}$$

as required. In Appendix A, as an illustration of this construction, we give a formula for S in the noninteracting ($\lambda=0$) case.

Up to this point the normalization of ψ_1 and ψ_2 has not been specified, and it is obviously immaterial for the construction of Eq. (4.5). However, for future use we now standardize the normalization by requiring that

$$\begin{aligned}&\frac{\partial\psi_1}{\partial\lambda}(\psi_1^\dagger\psi_1)^{-1/2}\to 0\ \text{ as }\theta\to 0\,,\\ &\frac{\partial\psi_2}{\partial\lambda}(\psi_2^\dagger\psi_2)^{-1/2}\to 0\ \text{ as }\theta\to\tfrac{1}{2}\pi\,,\end{aligned}\tag{4.7}$$

conditions which, as we shall see explicitly below, can be satisfied by taking the leading terms in the series developments of ψ_1 (ψ_2) about $\theta=0$ ($\theta=\frac{1}{2}\pi$) to be λ-independent constants.[7] Equation (4.7) uniquely specifies the λ dependence of ψ_1, ψ_2,

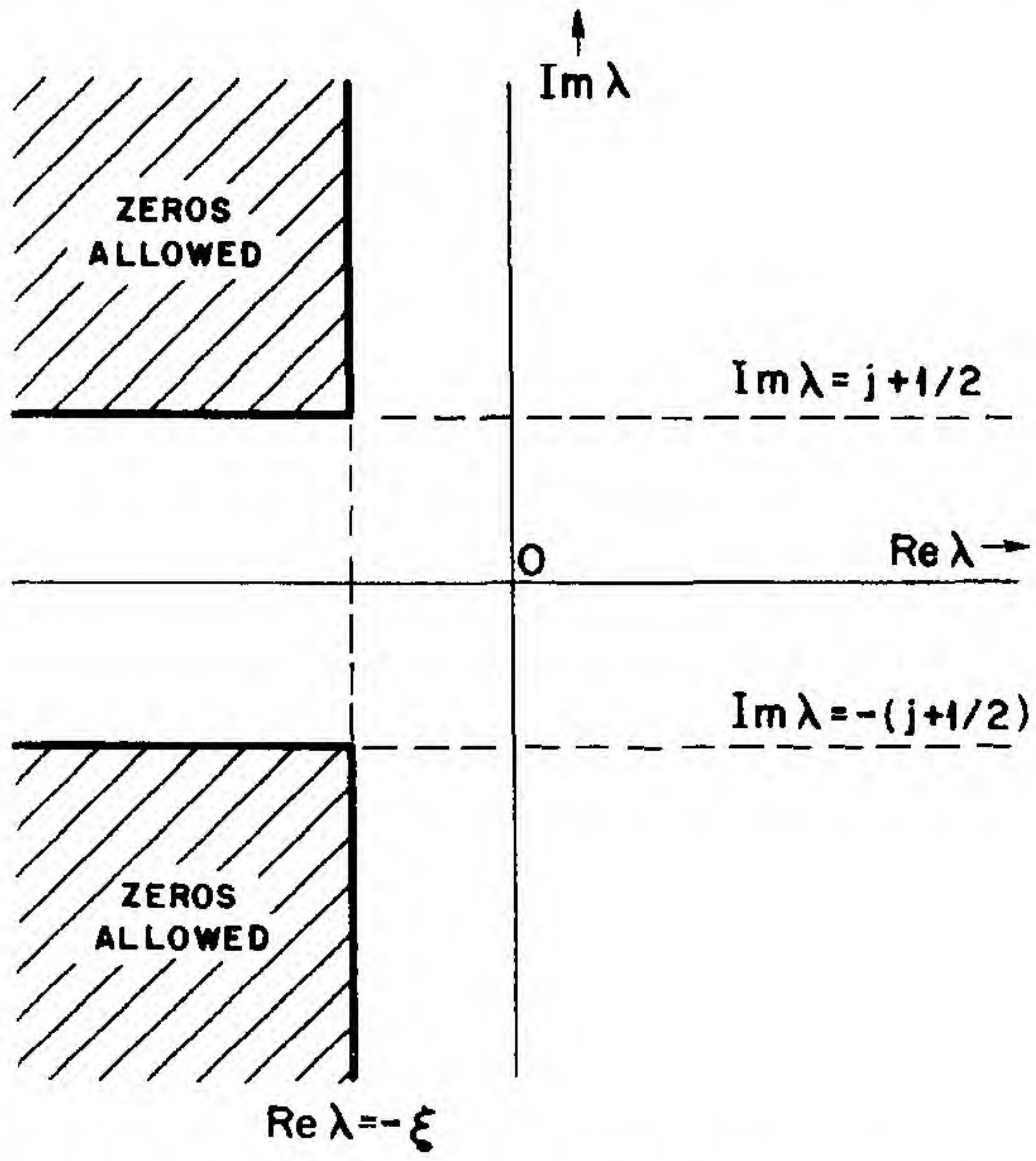


FIG. 1. Regions in which $\Delta_{\xi j}(\lambda)$ can have zeros according to the inequalities of Eqs. (3.9) and (3.12). We assume $\xi>0$.

and w, leaving arbitrary only a λ-independent normalization factor. It is now possible, by straightforward majorization arguments, to prove the following result: The standardized solutions $\psi_{1,2}$ are entire functions of λ, bounded for large λ by $e^{c|\lambda|}$, with c an appropriate constant.

B. Proof of the connection

To connect the Fredholm determinant $\Delta_{\xi j}(\lambda)$ with the Wronskian, we start from the formal relation[8]

$$\begin{aligned}\ln\Delta_{\xi j}(\lambda) &= \mathrm{Tr}\ln H\\ &= \mathrm{Tr}\ln\Big\{-[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\tau_1\\ &\qquad -(j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3 - i\frac{d}{d\theta}\tau_2\Big\},\end{aligned} \tag{4.8}$$

from which we get by differentiation

$$\Delta_{\xi j}(\lambda)^{-1}\frac{d\Delta_{\xi j}(\lambda)}{d\lambda} = \mathrm{Tr}\left[\frac{\partial H}{\partial\lambda}H^{-1}\right]. \tag{4.9}$$

Substituting Eq. (4.5) for $S=H^{-1}$ and evaluating the trace, we find

$$\Delta_{\xi j}(\lambda)^{-1}\frac{d\Delta_{\xi j}(\lambda)}{d\lambda} = w(\lambda)^{-1}\int_0^{\pi/2} d\theta\,\psi_2^T\frac{\partial H}{\partial\lambda}\psi_1\,. \tag{4.10}$$

We next show that the numerator on the right-hand side of Eq. (4.10) is just equal to $dw(\lambda)/d\lambda$ when ψ_1 and ψ_2 are taken to be the standard solutions. To see this, we start from Eq. (4.3) for w, which yields

$$\frac{dw(\lambda)}{d\lambda} = \frac{\partial\psi_2^T}{\partial\lambda}i\tau_2\psi_1 + \psi_2^T i\tau_2\frac{\partial\psi_1}{\partial\lambda}\,. \tag{4.11}$$

Letting $\theta\to\frac{1}{2}\pi$ and using Eq. (4.7), only the second term on the right-hand side of Eq. (4.11) survives, giving

$$\frac{dw(\lambda)}{d\lambda} = \psi_2^T i\tau_2\frac{\partial\psi_1}{\partial\lambda}\bigg|_{\theta\to\pi/2}\,. \tag{4.12}$$

To proceed we consider the integral appearing in the numerator of Eq. (4.10),

$$\int_0^{\pi/2} d\theta\,\psi_2^T\frac{\partial H}{\partial\lambda}\psi_1 = \lim_{\substack{\theta_1\to 0\\ \theta_2\to\pi/2}}\int_{\theta_1}^{\theta_2} d\theta\,\psi_2^T\frac{\partial H}{\partial\lambda}\psi_1\,. \tag{4.13}$$

By differentiating the equation $H\psi_1=0$ with respect to λ we get

$$\frac{\partial H}{\partial\lambda}\psi_1 = -H\frac{\partial\psi_1}{\partial\lambda} \tag{4.14}$$

and substituting this into Eq. (4.13), using the explicit form of H and integrating by parts, we find

$$\begin{aligned}\int_0^{\pi/2} d\theta\,\psi_2^T\frac{\partial H}{\partial\lambda}\psi_1 &= \lim_{\substack{\theta_1\to 0\\ \theta_2\to\pi/2}} -\int_{\theta_1}^{\theta_2} d\theta\,\psi_2^T H\frac{\partial\psi_1}{\partial\lambda}\\ &= \lim_{\substack{\theta_1\to 0\\ \theta_2\to\pi/2}} -\int_{\theta_1}^{\theta_2} d\theta\,\psi_2^T\left\{-i\frac{d}{d\theta}\tau_2 - [\xi(\sin\theta)^{-1}+\lambda\sin\theta]\tau_1 - (j+\tfrac{1}{2})(\cos\theta)^{-1}\tau_3\right\}\frac{\partial\psi_1}{\partial\lambda}\\ &= \lim_{\substack{\theta_1\to 0\\ \theta_2\to\pi/2}}\left[i\psi_2^T\tau_2\frac{\partial\psi_1}{\partial\lambda}\bigg|_{\theta_1}^{\theta_2} - \int_{\theta_1}^{\theta_2} d\theta (H\psi_2)^T\psi_1\right]\\ &= i\psi_2^T\tau_2\frac{\partial\psi_1}{\partial\lambda}\bigg|_{\theta\to\pi/2}\\ &= \frac{dw(\lambda)}{d\lambda}\,,\end{aligned} \tag{4.15}$$

giving the desired result. Substituting Eq. (4.15) into Eq. (4.10) we get, finally,

$$\Delta_{\xi j}(\lambda)^{-1}\frac{d\Delta_{\xi j}(\lambda)}{d\lambda} = w(\lambda)^{-1}\frac{dw(\lambda)}{d\lambda}\,, \tag{4.16}$$

which on integration gives the connection between the Fredholm determinant and the Wronskian,

$$\frac{\Delta_{\xi j}(\lambda)}{\Delta_{\xi j}(0)} = \frac{w(\lambda)}{w(0)}\,. \tag{4.17}$$

Since ψ_1, ψ_2 are entire functions of λ, we conclude that $\Delta_{\xi j}(\lambda)$ is also entire, as expected for a Fredholm determinant. Obviously, $\Delta_{\xi j}(\lambda)$ will also have exponentially bounded growth at infinity; the precise asymptotic form of $\Delta_{\xi j}(\lambda)$ will be given below. Equation (4.17) will be of great utility in the subsequent sections, where it will allow us to study $\Delta_{\xi j}(\lambda)$ by applying WKB and series expansion methods to the solutions of Eq. (3.7).

V. ORDER OF GROWTH OF $\Delta_{\xi j}(\lambda)$ AND $\bar{\Delta}[A]$

In this section we give more precise results concerning the large-λ asymptotic behavior of

$\Delta_{\xi j}(\lambda)$ and of the full external-field-problem Fredholm determinant $\Delta[A]$. In Sec. V A we present a WKB formula (derived in Appendix B) giving the asymptotic behavior of $\Delta_{\xi j}(\lambda)$. Using this formula, we determine the asymptotic distribution of zeros of $\Delta_{\xi j}(\lambda)$. In Sec. V B we discuss the renormalization subtractions needed to make the infinite product for $\Delta[A]$ convergent. Using our knowledge of the distribution of zeros of Δ, combined with results from the theory of entire functions, we determine the order of growth of the renormalized determinant $\bar{\Delta}[A]$ for large external-field amplitude λ. Combining this estimate with the absence of zeros in a strip containing the real axis, we show that the real amplitude integral contour discussed in Sec. I yields a function of e^2 analytic in the right-hand e^2 half plane.

A. Asymptotic behavior of $\Delta_{\xi j}(\lambda)$

As we have seen above, $\Delta_{\xi j}(\lambda)$ is given by the Wronskian of two suitably standardized independent solutions of the differential equation $H\psi=0$. In the limit when $|\lambda|$ is large, or more specifically, when the inequalities

$$\epsilon_1=\frac{\xi}{|\lambda|}\ll 1\,,\qquad \epsilon_2=\frac{j}{|\lambda|}\ll 1 \tag{5.1}$$

are satisfied, we can apply WKB methods to calculate approximate solutions of the differential equations, and hence to get the asymptotic form of $\Delta_{\xi j}(\lambda)$. The calculation, which is outlined in Appendix B, gives the result (valid for $\xi>0$)

$$\frac{\Delta_{\xi j}(\lambda)}{\Delta_{\xi j}(0)}\underset{\substack{\epsilon_1\ll 1\\ \epsilon_2\ll 1}}{\approx}\frac{\Gamma(2(j+1))}{\Gamma(j+\frac{3}{2})}\,\frac{\Gamma(\frac{1}{2})}{\Gamma(j+1)}\,\frac{\Gamma(j+\xi+1)}{\Gamma(\xi+\frac{1}{2})}\,2^{-2(j+1/2)}\lambda^{-(j+1/2)}\times[e^{\lambda}+e^{-\lambda}(-1)^{j+\xi+1}2^{-2(\xi+1/2)}(j+\tfrac{1}{2})\Gamma(\xi+\tfrac{1}{2})\lambda^{-(\xi+1/2)}]\,, \tag{5.2}$$

showing that the entire function $\Delta_{\xi j}(\lambda)$ is of exponential type. One special case of Eq. (5.2) is worth noting. When $j\to-\frac{1}{2}$, Eq. (5.2) reduces to

$$\left.\frac{\Delta_{\xi j}(\lambda)}{\Delta_{\xi j}(0)}\right|_{j\to-1/2}\underset{\substack{\epsilon_1\ll 1\\ \epsilon_2\ll 1}}{\approx}e^{\lambda}\,; \tag{5.3}$$

we will show in Sec. VI A below that this is an exact, and not just an asymptotic, result. From Eq. (5.2), we can calculate the asymptotic distribution of zeros of $\Delta_{\xi j}(\lambda)$ by solving the equation

$$0=e^{\lambda}+e^{-\lambda}(-1)^{j+\xi+1}2^{-2(\xi+1/2)}(j+\tfrac{1}{2})\times\Gamma(\xi+\tfrac{1}{2})\lambda^{-(\xi+1/2)}\,, \tag{5.4a}$$

which we rewrite in the form

$$\begin{aligned} &e^{2\lambda}(-\lambda)^{c_1}=e^{c_2}\\ &c_1=\xi+\tfrac{1}{2}\,,\\ &c_2=-2(\xi+\tfrac{1}{2})\ln 2+\ln(j+\tfrac{1}{2})\\ &\qquad+\ln\Gamma(\xi+\tfrac{1}{2})-i\pi(j+\tfrac{3}{2})\,. \end{aligned} \tag{5.4b}$$

Neglecting terms which vanish for large λ, the solution is

$$\begin{aligned} \lambda&\approx\pi ni+\tfrac{1}{2}c_2-\tfrac{1}{2}c_1\ln(-\pi ni)\\ &\approx\pi ni+\tfrac{1}{2}(\xi+\tfrac{1}{2})\ln\left(\frac{\xi+\frac{1}{2}}{|n|}\right)+O(\xi,j)\,,\\ \mathrm{Re}\lambda&\approx\tfrac{1}{2}(\xi+\tfrac{1}{2})\ln\left(\frac{\xi+\frac{1}{2}}{|n|}\right)+O(\xi,j)\,,\\ \mathrm{Im}\lambda&\approx\pi n+O(\xi,j)\,. \end{aligned} \tag{5.5}$$

In the region of validity of Eq. (5.5), where $|n|\gg\xi$, we see that $\mathrm{Re}\lambda$ is asymptotically negative, as required by the inequality of Eq. (3.9). The occurrence of zeros in complex-conjugate pairs is also apparent from Eq. (5.5).

For application in Sec. V B, it is convenient to give the zeros of $\Delta_{\xi j}(\lambda)$ an index k which arranges them in order of increasing magnitude:

$$\begin{aligned} &\lambda_k^{\xi j}=\text{general zero of }\Delta_{\xi j}(\lambda)\,,\\ &|\lambda_1^{\xi j}|\le|\lambda_2^{\xi j}|\le|\lambda_3^{\xi j}|<\cdots\,. \end{aligned} \tag{5.6}$$

For large k the index defined this way can be identified (up to a factor of two, since the zeros occur in complex-conjugate pairs) with the positive integer $|n|$ appearing in Eq. (5.5). Since the effective expansion parameters in the WKB procedure are thus

$$\frac{\xi}{|\lambda_k^{\xi j}|}\sim\frac{\xi}{k}\,,\quad\frac{j}{|\lambda_k^{\xi j}|}\sim\frac{j}{k}\,, \tag{5.7}$$

we expect the following bounds on $|\lambda_k^{\xi j}|$ to hold uniformly in ξ and j:

$$A_1\le\frac{|\lambda_k^{\xi j}|}{(\pi^2k^2+\frac{1}{4}(\xi+\frac{1}{2})^2\{\ln[(\xi+\frac{1}{2})/k]\}^2)^{1/2}}\le A_2\,,\qquad k\ge k_0=C(j^2+\xi^2)^{1/2}\,; \tag{5.8a}$$

$$(j^2+\xi^2)^{1/2}\le|\lambda_k^{\xi j}|\le A_3(j^2+\xi^2)^{1/2}\,,\quad k\le k_0 \tag{5.8b}$$

for suitable constants $A_{1,2,3}$ and C. [Equation (5.8b) also incorporates the lower bounds of Eqs. (3.9) and (3.12).] We have not constructed a proof of Eq. (5.8), so these inequalities should be

considered a conjecture, suggested by the WKB analysis, on which some of the arguments of Sec. V B are based.

B. Order of growth of $\Delta[A]$

We are now ready to examine the asymptotic behavior of the full external-field-problem Fredholm determinant $\Delta[A]$, given by the product formula Eq. (3.6). First we must deal with the question of renormalization subtractions alluded to above. By dividing out $\Delta_{\xi j}(0)^2$ in Eq. (3.6), we have eliminated the most divergent vacuum diagram illustrated in Fig. 2(a). However, the second-order diagram shown in Fig. 2(b) is also divergent, and must be eliminated by a further subtraction. To do this, we write the small-λ expansion

$$\frac{\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)}{\Delta_{\xi j}(0)^2}=1+A_{\xi j}\lambda^2+O(\lambda^4)\,, \tag{5.9}$$

and then define the renormalized Fredholm determinant $\tilde{\Delta}[A]$ by writing

$$\tilde{\Delta}[A]=e^{Q(\lambda)}\prod_{j=1/2,3/2,\ldots}\;\prod_{\xi=1/2,3/2,\ldots}\left[\frac{\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)}{\Delta_{\xi j}(0)^2}\,e^{-A_{\xi j}\lambda^2}\right]^{2j+1}. \tag{5.10}$$

In this expression

$$Q(\lambda)=Q_0+Q_2\lambda^2 \tag{5.11}$$

is a polynomial which expresses the fact that the renormalization counterterms always have an undetermined finite part. To see that Eq. (5.10) is the correct recipe, we note that the renormalized vacuum amplitude, which according to Eq. (1.8) is proportional to

$$\ln\tilde{\Delta}[A]=Q(\lambda)+\ln\Delta[A]-\ln\Delta[0]-\lambda^2\frac{d}{d\lambda^2}\ln\Delta[A]\bigg|_{\lambda^2=0}\,, \tag{5.12}$$

now receives contributions only from the convergent vacuum diagrams illustrated in Fig. 3.

Let us next rewrite Eq. (5.10) in an alternative useful form. Since $\Delta_{\xi j}(\lambda)$ is an entire function of exponential type, we can use the Hadamard factorization theorem[9] to write it as an infinite product in terms of its zeros $\lambda_k^{\xi j}$,

$$\frac{\Delta_{\xi j}(\lambda)}{\Delta_{\xi j}(0)}=e^{\alpha_{\xi j}\lambda}\prod_k\left(1-\frac{\lambda}{\lambda_k^{\xi j}}\right)\exp\left(\frac{\lambda}{\lambda_k^{\xi j}}\right), \tag{5.13}$$

giving

$$\begin{aligned}\frac{\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)}{\Delta_{\xi j}(0)^2}&=\prod_k\left[1-\frac{\lambda^2}{(\lambda_k^{\xi j})^2}\right]\\&=1-\lambda^2\sum_k\frac{1}{(\lambda_k^{\xi j})^2}+O(\lambda^4)\,.\end{aligned} \tag{5.14}$$

From Eq. (5.14) we identify $A_{\xi j}$ as

$$A_{\xi j}=-\sum_k\frac{1}{(\lambda_k^{\xi j})^2}\,. \tag{5.15}$$

Let us define an additional constant $B_{\xi j}$ by

$$B_{\xi j}=\sum_k\frac{1}{(\lambda_k^{\xi j})^4} \tag{5.16}$$

and combine Eqs. (5.13)–(5.16) to rewrite Eq. (5.10) as

$$\begin{aligned}\tilde{\Delta}[A]&=e^{Q(\lambda)}e^{-B\lambda^4/2}P(\lambda)\,,\\P(\lambda)&=\prod_{j=1/2,3/2,\ldots}\;\prod_{\xi=1/2,3/2,\ldots}\;\prod_k\left\{\left[1-\frac{\lambda^2}{(\lambda_k^{\xi j})^2}\right]\exp\left[\left(\frac{\lambda}{\lambda_k^{\xi j}}\right)^2+\frac{1}{2}\left(\frac{\lambda}{\lambda_k^{\xi j}}\right)^4\right]\right\}^{2j+1}\\&=\prod_{\substack{\text{all zeros}\\ \lambda_\nu\text{ of }\tilde{\Delta}[A]}}\left\{\left(1-\frac{\lambda}{\lambda_\nu}\right)\exp\left[\frac{\lambda}{\lambda_\nu}+\frac{1}{2}\left(\frac{\lambda}{\lambda_\nu}\right)^2+\frac{1}{3}\left(\frac{\lambda}{\lambda_\nu}\right)^3+\frac{1}{4}\left(\frac{\lambda}{\lambda_\nu}\right)^4\right]\right\},\\B&=\sum_{j=1/2,3/2,\ldots}\;\sum_{\xi=1/2,3/2,\ldots}(2j+1)B_{\xi j}\,.\end{aligned} \tag{5.17}$$

The constant B is the contribution of the fourth-order graph which appears as the first term in the series of Fig. 3, and hence is finite. The second expression for $P(\lambda)$ in Eq. (5.17) has the form called a *canonical product* in the theory of entire functions[9]; Eq. (5.17) thus expresses $\tilde{\Delta}[A]$ as a canonical product multiplied by the exponential of a fourth-degree polynomial in λ.

(a) (b)

FIG. 2. (a) Divergent vacuum diagram which is removed by division by $\Delta_{\xi j}(0)^2$ in Eq. (3.6). (b) Divergent vacuum diagram which is removed by the factor $\exp(-A_{\xi j}\lambda^2)$ in Eq. (5.10).

Let us now introduce some further concepts from the theory of entire functions.[9] Let $f(\lambda)$ be an entire function of the complex variable λ. Its *maximum modulus* $M(r)$ and *minimum modulus* $m(r)$ are defined by

$$\begin{aligned} M(r) &= \max_{0\le\theta\le 2\pi} |f(re^{i\theta})| \,, \\ m(r) &= \min_{0\le\theta\le 2\pi} |f(re^{i\theta})| \,. \end{aligned} \tag{5.18}$$

The *order* ρ of $f(\lambda)$ is defined to be

$$\rho = \limsup_{r\to\infty} \frac{\ln\ln M(r)}{\ln r} \,; \tag{5.19a}$$

if f is of order ρ it is asymptotically bounded by

$$|f(\lambda)| \le A e^{B|\lambda|^{\rho}} \tag{5.19b}$$

for suitable positive constants A and B. Finally, let $\{r_\nu = |\lambda_\nu|\}$ be the sequence of moduli of the zeros λ_ν of $f(\lambda)$, arranged in increasing order. The smallest number σ for which

$$\sum_{\nu=1}^{\infty} \frac{1}{r_\nu{}^{\alpha}} < \infty, \quad \text{for all } \alpha > \sigma \tag{5.20}$$

is called the *exponent of convergence* of the sequence. According to the theory of entire functions, the order of an entire function is closely related to the exponent of convergence of its zeros.

To determine the order of $\Delta[A]$, we wish then to calculate the exponent of convergence of the zeros λ_ν appearing in Eq. (5.17). Remembering that all zeros $\lambda_k^{\xi j}$ occur with multiplicity $2j+1$, we consider the sum

$$S_\alpha = \sum_{j=1/2,3/2,\ldots} (2j+1) \sum_{\xi=1/2,3/2,\ldots} \sum_k \frac{1}{|\lambda_k^{\xi j}|^{\alpha}} \,. \tag{5.21}$$

In estimating the convergence properties of S_∞ it obviously suffices to replace the sums in Eq. (5.21) by integrals. We first show that Eq. (5.21) is convergent for $\alpha > 4$. Using the lower bounds obtained from Eq. (5.8),

$$\begin{aligned} \pi k A_1 &\le |\lambda_k^{\xi j}|, \quad k \ge k_0 \\ (j^2+\xi^2)^{1/2} &\le |\lambda_k^{\xi j}|, \quad k \le k_0 \end{aligned} \tag{5.22}$$

we get the estimate

$$\begin{aligned} \sum_k \frac{1}{|\lambda_k^{\xi j}|^{\alpha}} &= \sum_{k=1}^{k_0} \frac{1}{|\lambda_k^{\xi j}|^{\alpha}} + \sum_{k=k_0}^{\infty} \frac{1}{|\lambda_k^{\xi j}|^{\alpha}} \\ &\lesssim \frac{C(j^2+\xi^2)^{1/2}}{(j^2+\xi^2)^{\alpha/2}} + \int_{C(j^2+\xi^2)^{1/2}}^{\infty} \frac{dk}{(\pi k A_1)^{\alpha}} \\ &= \frac{C'}{(j^2+\xi^2)^{(\alpha-1)/2}} \,, \end{aligned} \tag{5.23}$$

so that

FIG. 3. Convergent diagrams which contribute to Eq. (5.10).

$$S_\alpha \lesssim \int_{1/2}^{\infty} d\xi \int_{1/2}^{\infty} j dj \, \frac{4C'}{(j^2+\xi^2)^{(\alpha-1)/2}} < \frac{4C' 2^{\alpha-4}}{(\alpha-3)(\alpha-4)} < \infty \,, \tag{5.24}$$

as claimed. Next we show that Eq. (5.21) diverges when $\alpha \to 4$. Since $(\ln x)/x \le 1/e$ for $x \ge 1$, the upper bounds in Eq. (5.8) take the form

$$\begin{aligned} |\lambda_k^{\xi j}| &\le A_2 \pi \left(1 + \frac{1}{4e^2}\right)^{1/2} k, \quad k \ge k_0 \\ |\lambda_k^{\xi j}| &\le A_3 (j^2+\xi^2)^{1/2}, \quad k \le k_0 \end{aligned} \tag{5.25}$$

giving, by a procedure identical to that in Eqs. (5.23) and (5.24), the estimate

$$S_\alpha \gtrsim \frac{C''}{(\alpha-4)} \,, \quad C'' > 0 \,. \tag{5.26}$$

We conclude that S_α diverges for $\alpha = 4$, and that the exponent of convergence of the zeros of $\Delta[A]$ is $\sigma = 4$.

From the fact that $\sigma = 4$ we can immediately conclude that the order of the canonical product $P(\lambda)$ is 4, and hence that the order of $\Delta[A]$ is less than or equal to 4.[9] If the order of $\Delta[A]$ were actually less than 4, then the sum in Eq. (5.21) would converge[9] for exponents α smaller than 4, which we have seen is not the case. So we conclude that the order of $\Delta[A]$ is precisely 4.

Let us now use these results to determine the convergence properties of the amplitude integral when taken along the real contour. Since $\Delta_{\xi j}(\lambda)$ cannot change sign on the real axis, all of the factors in Eq. (5.10), and hence $\Delta[A]$ itself, are positive for λ real, and so $\ln\bar{\Delta}[A]$ is real. Since the maximum modulus of $\bar{\Delta}[A]$ is bounded as in Eq. (5.19b) with $\rho = 4$, we have

$$\ln\Delta[A] < B|\lambda|^4 \tag{5.27}$$

for an appropriate positive constant B. In order to restrict $\ln\Delta[A]$ from below, it is necessary to have a lower bound on the minimum modulus of $\Delta[A]$. We get this by using the following theorem[9]: "Let $P(\lambda)$ be a canonical product of order ρ. About each zero λ_ν ($|\lambda_\nu| > 1$) we draw a circle of radius $1/|\lambda_\nu|^{\alpha}$, $\alpha > \rho$. Then in the region outside these excluded circles, $|P(\lambda)| > \exp(-r^{\rho+\epsilon})$ for $\epsilon > 0$ and for $r > r_0(\epsilon, \alpha)$." To apply this theorem, we note that the sum of the radii of all the circles is just S_∞ and can be made smaller than 1 by choosing α large enough. Since $\Delta[A]$ has no zeros in the strip $|\text{Im}\lambda| \le 1$, the entire real axis then lies

in the region outside the excluded circles, and so we learn

$$\ln\Delta[A] > -|\lambda|^{4+\epsilon}, \quad |\lambda| > r_0 \tag{5.28}$$

for r_0 appropriately large. Taking Eqs. (5.27) and (5.28) together, we see that $|\ln\Delta[A]|$ is polynomial-bounded on the λ-real axis. The Gaussian factor in Eq. (1.5) then guarantees that the amplitude integral converges when taken along the real axis, provided that $\mathrm{Re}e^2 > 0$, and thus defines a function of e^2 analytic in the right-hand e^2 half plane. Note that this conclusion does not depend on the fact that $\Delta[A]$ is of order 4, but only requires the weaker statement that the order of $\Delta[A]$ is finite, which is known to be true[2] independent of the validity of the inequalities in Eq. (5.8).

VI. NUMERICAL RESULTS

We turn next to numerical studies of $\Delta_{\xi j}(\lambda)$ and $\tilde{\Delta}[A]$. In Sec. VIA we derive power-series expansions for the standardized solutions ψ_1 and ψ_2. The circles of convergence of the two series which we obtain overlap, allowing one to compute the Wronskian, and hence $\Delta_{\xi j}(\lambda)$, by picking θ to have any value in the overlap region. In Sec. VIB we numerically study the location of low-lying zeros of $\Delta_{\xi j}(\lambda)$, and find that there are no zeros in the sectors $|\mathrm{Re}\lambda| > |\mathrm{Im}\lambda|$. Consequences of this fact for the coupling-constant analyticity properties of W_1 are discussed. Finally, in Sec. VIC we give numerical results for the behavior of the vacuum amplitude as λ increases along the imaginary axis.

A. Power-series solutions

Substituting

$$\psi = \begin{pmatrix} a \\ c \end{pmatrix} \tag{6.1}$$

into Eq. (3.7) and writing out the coupled differential equations for the two components, we get

$$\frac{da}{d\theta} - [\xi(\sin\theta)^{-1} + \lambda\sin\theta]a + (j+\tfrac{1}{2})(\cos\theta)^{-1}c = 0,$$
$$\tag{6.2}$$
$$\frac{dc}{d\theta} + [\xi(\sin\theta)^{-1} + \lambda\sin\theta]c + (j+\tfrac{1}{2})(\cos\theta)^{-1}a = 0.$$

To construct power-series solutions regular around $\theta=0$ and $\theta=\frac{1}{2}\pi$ we make the following changes of variable, motivated by the form of the noninteracting ($\lambda=0$) solutions presented in Appendix A.

(1) Solution ψ_1 regular around $\theta=0$. We substitute

$$a_1 = (\tan\tfrac{1}{2}\theta)^{\xi} f(x),$$
$$c_1 = \tan\theta(\tan\tfrac{1}{2}\theta)^{\xi} g(x), \tag{6.3}$$
$$x = 1 - \frac{1}{\cos\theta}.$$

In terms of the new variables the coupled equations become

$$\frac{df}{dx} + \frac{\lambda f}{(1-x)^2} - (j+\tfrac{1}{2})g = 0,$$
$$\tag{6.4}$$
$$x(2-x)\left[\frac{dg}{dx} - \frac{\lambda g}{(1-x)^2}\right] + (2\xi+1-x)g + (j+\tfrac{1}{2})f = 0.$$

We now look for a power-series solution in the form

$$f = \sum_{n=0}^{\infty} x^n f_n, \quad g = \sum_{n=0}^{\infty} x^n g_n. \tag{6.5}$$

We find that Eqs. (6.4) are satisfied if we take

$$f_n = g_n = 0 \quad (n<0), \quad f_0 = -2(\xi+\tfrac{1}{2}), \quad g_0 = (j+\tfrac{1}{2}),$$
$$f_{n+1} = \frac{1}{n+1}[(2n-\lambda)f_n - (n-1)f_{n-1} + (j+\tfrac{1}{2})(g_n - 2g_{n-1} + g_{n-2})]$$
$$\tag{6.6}$$
$$g_{n+1} = \frac{1}{2n+2\xi+3}[(5n+4\xi+3+2\lambda)g_n - (4n+2\xi-1+\lambda)g_{n-1} + (n-1)g_{n-2} - (j+\tfrac{1}{2})(f_{n+1} - 2f_n + f_{n-1})], \quad n \geq 0.$$

(2) Solution ψ_2 regular around $\theta=\frac{1}{2}\pi$. In this case we make the substitution

$$a_2 = \left(\frac{\cos\theta}{1+\sin\theta}\right)^{j+1/2}[h(y) + \cos\theta\, l(y)],$$
$$c_2 = \left(\frac{\cos\theta}{1+\sin\theta}\right)^{j+1/2}[h(y) - \cot\theta\, l(y)], \tag{6.7}$$
$$y = 1 - \frac{1}{\sin\theta}.$$

The coupled differential equations now become

$$\frac{dh}{dy} - \xi l - \frac{\lambda}{(1-y)^2} l = 0,$$
$$\tag{6.8}$$
$$y(2-y)\frac{dl}{dy} + [2(j+1) - y]l + \left[\xi + \frac{\lambda}{(1-y)^2}\right]h = 0.$$

Assuming power-series solutions in the form

$$h = \sum_{n=0}^{\infty} h_n y^n, \quad l = \sum_{n=0}^{\infty} l_n y^n, \tag{6.9}$$

we find the solutions

$$h_n = l_n = 0 \quad (n<0), \quad h_0 = -2(j+1), \quad l_0 = \xi + \lambda,$$

$$h_{n+1} = \frac{1}{n+1}[2nh_n - (n-1)h_{n-1} + (\xi+\lambda)l_n + \xi(l_{n-2} - 2l_{n-1})], \tag{6.10}$$

$$l_{n+1} = \frac{1}{2n+2j+4}[(5n+4j+5)l_n - (4n+2j)l_{n-1} + (n-1)l_{n-2} - (\lambda+\xi)h_{n+1} + \xi(2h_n - h_{n-1})], \quad n \geq 0.$$

A number of observations about the above solutions are now in order. First, we note that since

$$\frac{\partial f_0}{\partial\lambda} = \frac{\partial g_0}{\partial\lambda} = \frac{\partial h_0}{\partial\lambda} = 0, \tag{6.11}$$

and since l in Eq. (6.7) appears multiplied by the factor $\cot\theta$, which vanishes at $\theta = \frac{1}{2}\pi$, the standardization conditions of Eq. (4.7) are satisfied. Second, we consider the greatly simplified form of the above equations when $j \to -\frac{1}{2}$. Working directly from Eq. (6.2) we find in this special limit the decoupled equations

$$\begin{aligned} &\frac{da}{d\theta} - [\xi(\sin\theta)^{-1} + \lambda\sin\theta]a = 0, \\ &\frac{dc}{d\theta} + [\xi(\sin\theta)^{-1} + \lambda\sin\theta]c = 0, \end{aligned} \tag{6.12}$$

which can be immediately integrated, giving

$$\begin{aligned} a_1 &= -2(\tan\tfrac{1}{2}\theta)^{\xi}(\xi+\tfrac{1}{2})e^{\lambda(1-\cos\theta)}, \\ c_1 &= 0, \\ a_2 &= -(\tan\tfrac{1}{2}\theta)^{\xi} e^{-\lambda\cos\theta}, \\ c_2 &= -(\tan\tfrac{1}{2}\theta)^{-\xi} e^{\lambda\cos\theta}. \end{aligned} \tag{6.13}$$

Hence the Wronskian is

$$\begin{aligned} w(\lambda) &= a_2 c_1 - a_1 c_2 \\ &= -2(\xi+\tfrac{1}{2})e^{\lambda}, \end{aligned} \tag{6.14}$$

giving for the $j \to -\frac{1}{2}$ limit of the Fredholm determinant the result

$$\frac{\Delta_{\xi\,-1/2}(\lambda)}{\Delta_{\xi\,-1/2}(0)} = e^{\lambda}, \tag{6.15}$$

as was stated in Sec. V A above.

Finally, we discuss the convergence properties of the power-series solutions. Rewriting Eq. (6.4) as a single second-order differential equation we find singular points at $x=1$, 2, and ∞. Rewriting Eq. (6.8) as a single second-order equation we find singular points at $y=1$, 2, and ∞, and additionally at

$$y = 1 \pm \left(\frac{-\lambda}{\xi}\right)^{1/2}. \tag{6.16}$$

Since x and y are related by

$$\frac{1}{(1-x)^2} + \frac{1}{(1-y)^2} = 1, \tag{6.17}$$

Eq. (6.16) corresponds to singular points in the x variable at

$$x = 1 \pm \left(1 + \frac{\xi}{\lambda}\right)^{-1/2}, \tag{6.18}$$

which did not appear in the x form of the equation. Hence the singularities in Eq. (6.16) must be removable, and a direct calculation shows this to be the case. We conclude, then, that the power-series solutions for ψ_1 and ψ_2 have the following regions of convergence:

$$\begin{aligned} &\psi_1 \text{ converges for } |x|<1 \Leftrightarrow 1 \geq \cos\theta > \tfrac{1}{2} \Leftrightarrow 0 \leq \theta < \tfrac{1}{3}\pi, \\ &\psi_2 \text{ converges for } |y|<1 \Leftrightarrow 1 \geq \sin\theta > \tfrac{1}{2} \Leftrightarrow \tfrac{1}{6}\pi < \theta \leq \tfrac{1}{2}\pi. \end{aligned} \tag{6.19}$$

Thus, in the angular range $\frac{1}{6}\pi < \theta < \frac{1}{3}\pi$ both power series are convergent, and so we can calculate the Wronskian from Eq. (4.3) by taking θ to be any value in this interval. Since the Wronskian is θ-independent, a powerful check on both the programming and the absence of serious round-off and truncation errors is obtained by calculating W for two different values of θ in the allowed range and then checking that the same answer is obtained. In practice, using double precision on an IBM 360/91, we found we were able to explore the region $\xi \leq 80$, $j \leq 80$, $|\lambda| \leq 20$ in good detail, but for $|\lambda|$ values between 20 and 24, serious roundoff errors started to set in.

B. Low-lying zeros of $\Delta_{\xi j}(\lambda)$

Numerical results for the low-lying zeros of $\Delta_{\xi j}(\lambda)$ in the upper half plane are given in Tables II and III. In Table II we give the locations of the lowest zero (the zero of smallest magnitude $|\lambda|$) for a range of values of ξ and j. In Table III we give the locations of the lowest four zeros for $\xi = j = \frac{1}{2}$. For all of the zeros tabulated, the ratio $|\mathrm{Im}\lambda|/|\mathrm{Re}\lambda|$ is larger than 1. As ξ increases for fixed j, the ratio appears to be approaching 1 from above; as j increases for fixed ξ, the ratio grows, as might be expected from the inequality of Eq. (3.12). For a given ξ, j, the successive higher zeros move up in the imaginary direction with a spacing $\sim\pi$ between the imaginary parts, as is expected from the WKB estimate of Eq. (5.5). The pattern of the numerical results strongly suggests that $|\mathrm{Im}\lambda|/|\mathrm{Re}\lambda| > 1$ for all zeros of

$\Delta_{\xi j}(\lambda)$. If this property were true, the zero-free regions of $\tilde{\Delta}[A]$ would be as indicated in Fig. 4, and a contour of integration in Eq. (1.5) initially along the real axis could be freely deformed to the positions indicated as "# 1" and "# 2." The first (second) contour allows analytic continuation of W_1 into the entire upper (lower) e^2 half plane. Hence, for the distribution of zeros of $\tilde{\Delta}[A]$ shown in Fig. 4 one gets a radiative-corrected vacuum amplitude W_1 which is analytic in the entire e^2 plane except for a branch cut running along the negative real axis from 0 to $-\infty$.

C. Behavior of vacuum amplitude for λ imaginary

As we have stressed repeatedly above, the possibility of taking the contour in Eq. (1.5) to lie along the imaginary axis can be realized only if $W^{(0)}$ decreases as a Gaussian (or faster) as λ becomes infinite along the imaginary axis. Actually, when subtractions are taken into account, the relevant question becomes whether $(d/d\lambda^2)^2 \ln\tilde{\Delta}[A]$ decreases along the imaginary axis. The differentiations just eliminate the arbitrary subtraction polynomial $Q(\lambda)$ which appears in Eq. (5.10); this polynomial is not relevant to the physics, and specifically is not present if we consider (in the one-mode approximation for virtual photons) the set of single-fermion-loop vacuum polarization diagrams shown in Fig. 5. In order to obtain good convergence of the sum over separation parameters ξ, j, we found it necessary to differentiate once more with respect to λ^2. Multiplying (for convenience) by λ^2, we get, finally, as the quantity being studied

$$\tilde{W}^{(0)}(\lambda) \equiv \lambda \frac{d}{d\lambda}\left(\frac{1}{\lambda}\frac{d}{d\lambda}\right)^2 \ln\tilde{\Delta}[A]. \tag{6.20}$$

Results for $\tilde{W}^{(0)}$ versus $-i\lambda$ are shown in Fig. 6. In calculating the points for this curve, we summed on ξ from $\frac{1}{2}$ to $2\frac{1}{2}$ and on j from $\frac{1}{2}$ to $39\frac{1}{2}$; doubling both summation ranges for a subset of the points produced a 6% change for $-i\lambda=1$ and negligible (<1%) change for $-i\lambda \geq 5$. In fact, nearly all of the sum for $-i\lambda \geq 5$ came from $\Delta_{\xi j}$'s with $\xi=\frac{1}{2}$, most likely a result of the fact that this is the value of ξ which gives zeros of $\Delta_{\xi j}$ lying closest to the imaginary axis (see Table II). The curve plotted shows no sign of a rapid decrease, but unfortunately the distortions in both the envelope of the oscillations and the wave form suggest that the asymptotic region has not been reached, and so the results are inconclusive. We did not attempt to extend the computations further, because of the roundoff error problem mentioned above.

TABLE II. Lowest-lying zero with $\mathrm{Im}\lambda > 0$ for various ξ, j values.

j	ξ	$\mathrm{Re}\lambda_1$	$\mathrm{Im}\lambda_1$	$\lvert\mathrm{Im}\lambda_1\rvert/\lvert\mathrm{Re}\lambda_1\rvert$
$\frac{1}{2}$	$\frac{1}{2}$	−1.67	7.12	4.26
$\frac{1}{2}$	$\frac{3}{2}$	−3.47	7.36	2.12
$\frac{1}{2}$	$\frac{7}{2}$	−6.46	8.50	1.32
$\frac{1}{2}$	$\frac{11}{2}$	−9.87	12.57	1.27
$\frac{1}{2}$	$\frac{15}{2}$	−13.25	16.65	1.26
$\frac{1}{2}$	$\frac{1}{2}$	−1.67	7.12	4.26
$\frac{3}{2}$	$\frac{1}{2}$	−1.43	8.93	6.24
$\frac{5}{2}$	$\frac{1}{2}$	−1.14	7.73	6.78
$\frac{7}{2}$	$\frac{1}{2}$	−1.10	9.71	8.83
$\frac{9}{2}$	$\frac{1}{2}$	−1.08	11.70	10.83
$\frac{11}{2}$	$\frac{1}{2}$	−1.17	16.60	14.19
$\frac{13}{2}$	$\frac{1}{2}$	−1.14	18.59	16.31
$\frac{1}{2}$	$\frac{1}{2}$	−1.67	7.12	4.26
$\frac{3}{2}$	$\frac{3}{2}$	−2.94	6.27	2.13
$\frac{7}{2}$	$\frac{7}{2}$	−6.13	10.98	1.79
$\frac{11}{2}$	$\frac{11}{2}$	−9.86	18.67	1.89

VII. OPEN QUESTIONS

We conclude by giving a brief recapitulation of the remaining unresolved questions. Within the framework of the one-mode approximation discussed at great length above, some key problems are:

(i) determining the asymptotic behavior of $\tilde{W}^{(0)}(\lambda)$ along the imaginary axis (ruling out a Gaussian decrease would rule out the imaginary contour possibility and hence, as discussed in Ref. 2, would rule out the possibility of obtaining a coupling-constant eigenvalue when only a finite number of photon modes are included),

(ii) proving (or disproving) the distribution of zeros illustrated in Fig. 4,

(iii) if Fig. 4 is correct, finding a simple formula or interpretation for the discontinuity of W_1

TABLE III. First four zeros for $\xi=j=\frac{1}{2}$ with $\mathrm{Im}\lambda>0$. (For each there is a corresponding complex-conjugate zero in the lower half plane.)

Zero number k	$\mathrm{Re}\lambda_k$	$\mathrm{Im}\lambda_k$	$\lvert\mathrm{Im}\lambda_k\rvert/\lvert\mathrm{Re}\lambda_k\rvert$	$\mathrm{Im}\lambda_k - \mathrm{Im}\lambda_{k-1}$
1	−1.67	7.12	4.26	
2	−1.86	10.23	5.50	3.11
3	−1.99	13.39	6.73	3.16
4	−2.10	16.52	7.87	3.13

across its cut in the e^2 plane, and

(iv) finding a compact expression for $\Delta_{\xi j}(\lambda)$ in which the parameter θ in the Wronskian has been explicitly eliminated.

Going beyond the one-mode problem to the case when a finite number of photon modes are present, one can ask whether the zero-free regions shown in Fig. 4 persist.[10,11] If so, then the real contour would give cut-plane analyticity in e^2 for any finite number of modes, and the important (and undoubtedly difficult) question of what happens when the limit to an infinite number of modes is taken would be brought to the fore.

ACKNOWLEDGMENTS

I wish to thank S. Coleman, S. B. Treiman, A. S. Wightman, and T. T. Wu for helpful conversations, and to acknowledge the hospitality of the Aspen Center for Physics and the National Accelerator Laboratory, where parts of this work were done.

APPENDIX A: FREE GREEN'S FUNCTION

We give here a closed-form expression for the Green's function of Eq. (4.5) in the free ($\lambda = 0$) case. The result is most compactly expressed in terms of the Jacobi functions

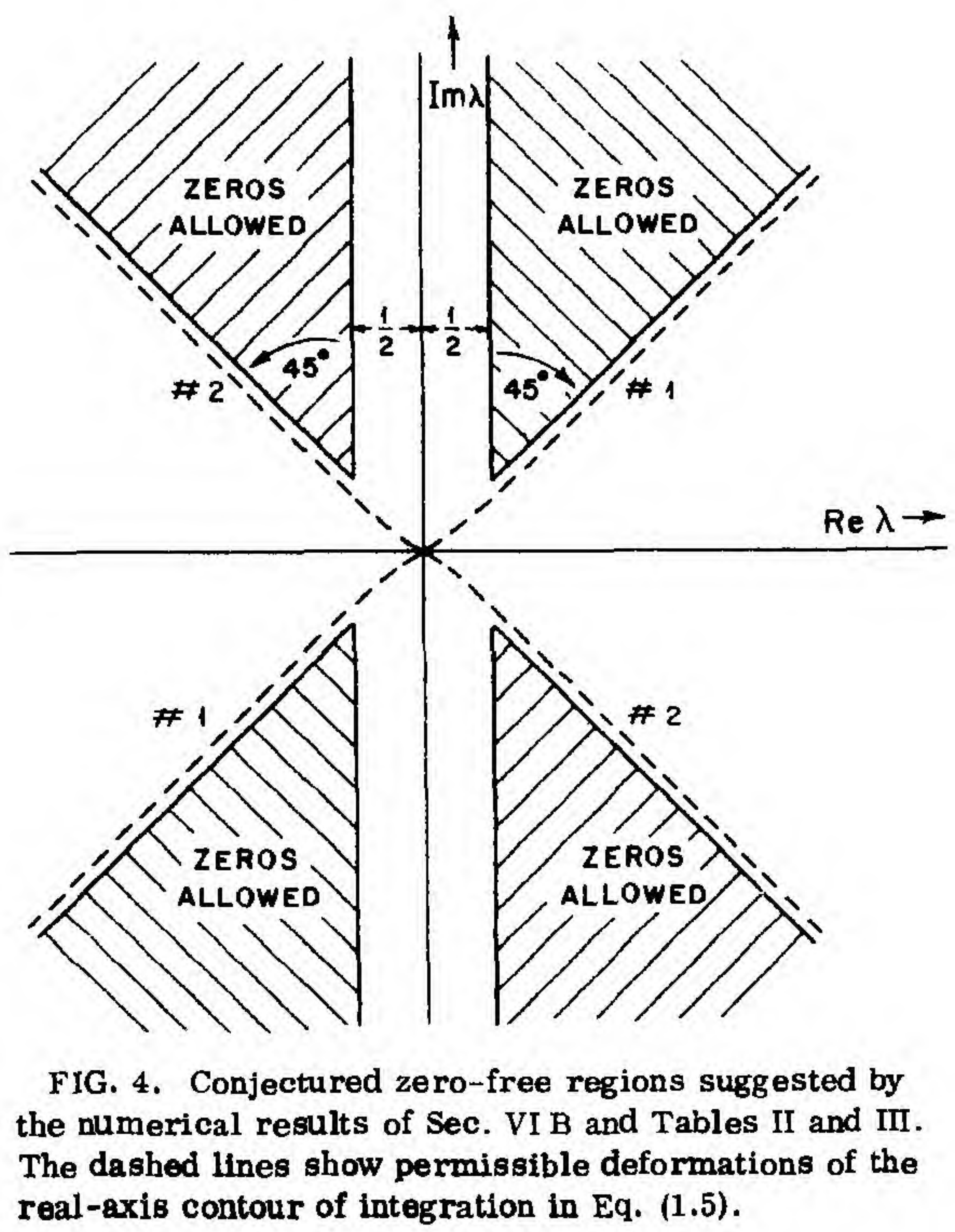


FIG. 4. Conjectured zero-free regions suggested by the numerical results of Sec. VI B and Tables II and III. The dashed lines show permissible deformations of the real-axis contour of integration in Eq. (1.5).

$$P_\nu^{(\alpha,\beta)}(z) = \frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)\,\Gamma(\alpha+1)} \times F(-\nu,\ \nu+\alpha+\beta+1;\ \alpha+1;\ \tfrac{1}{2}-\tfrac{1}{2}z), \tag{A1}$$

where $F(a,b;c;z)$ is the usual hypergeometric function. [The ordinary Jacobi polynomials correspond to the case where ν in Eq. (A1) is a nonnegative integer. We will also use the case where ν is a non-negative half-integer.] We find (for $\xi > 0$)

$$\psi_1^0 = \begin{pmatrix} a_1^0 \\ c_1^0 \end{pmatrix},$$

$$a_1^0 = (\tan\tfrac{1}{2}\theta)^{\xi}\, P_{j+1/2}^{(\xi-1/2,\,-\xi-1/2)}\left(\frac{1}{\cos\theta}\right),$$

$$c_1^0 = -\tfrac{1}{2}\tan\theta(\tan\tfrac{1}{2}\theta)^{\xi}\, P_{j-1/2}^{(\xi+1/2,\,-\xi+1/2)}\left(\frac{1}{\cos\theta}\right);$$

$$\psi_2^0 = \begin{pmatrix} a_2^0 \\ c_2^0 \end{pmatrix}, \tag{A2}$$

$$a_2^0 = \left(\frac{\cos\theta}{1+\sin\theta}\right)^{j+1/2}\left(1+\frac{\sin\theta}{\xi}\frac{d}{d\theta}\right) P_\xi^{(j,\,-j-1)}\left(\frac{1}{\sin\theta}\right),$$

$$c_2^0 = \left(\frac{\cos\theta}{1+\sin\theta}\right)^{j+1/2}\left(1-\frac{\sin\theta}{\xi}\frac{d}{d\theta}\right) P_\xi^{(j,\,-j-1)}\left(\frac{1}{\sin\theta}\right).$$

The Wronskian of the two solutions is easily calculated by taking either the limit $\theta \to 0$ or the limit $\theta \to \frac{1}{2}\pi$, giving

$$w^0 = a_2^0 c_1^0 - a_1^0 c_2^0 = \frac{-\Gamma(j+\xi+1)}{\Gamma(\frac{1}{2})\,\Gamma(\xi+1)\,\Gamma(j+\frac{3}{2})}. \tag{A3}$$

The free Green's function then immediately follows from the recipe of Eq. (4.5),

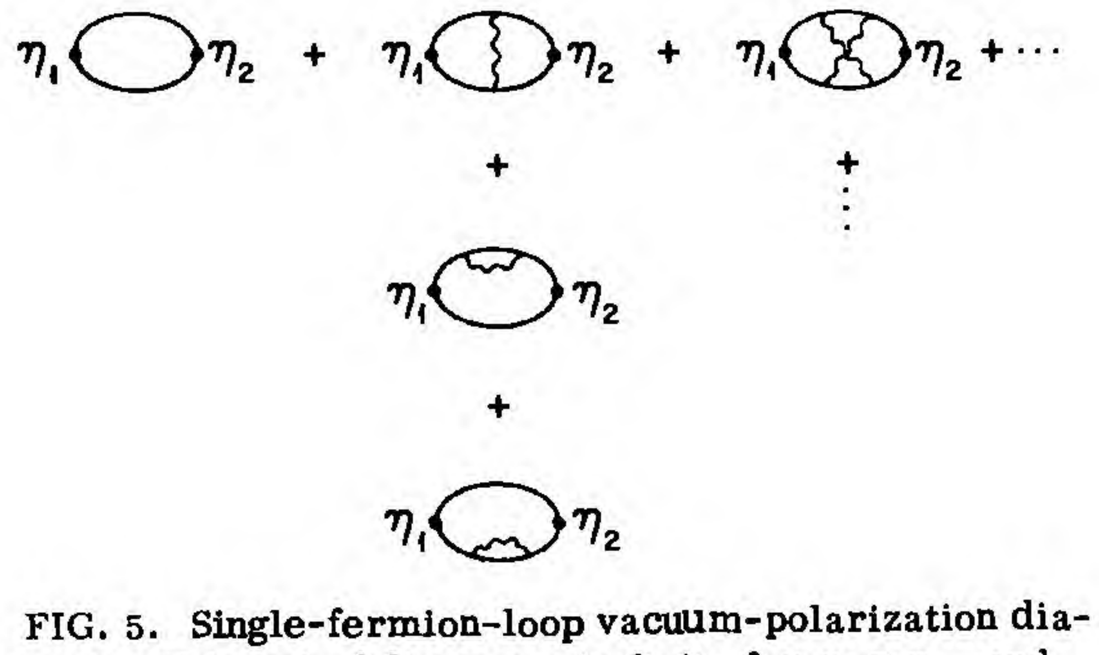


FIG. 5. Single-fermion-loop vacuum-polarization diagrams. This set of diagrams is finite for $\eta_1 \neq \eta_2$, and requires no subtractions. However, if we contract with $Y_{1Ma}^{(1)}(\eta_1)\, Y_{1Mb}^{(2)}(\eta_2)$ and integrate over η_1 and η_2, the short-distance singularity as $\eta_1 \to \eta_2$ leads to a divergence, corresponding to the $A_{\xi j}$ counterterm in Eq. (5.10) and the finite remainder $Q_2\lambda^2$ in Eq. (5.11). This divergence is of no physical significance, and so we differentiate to eliminate Q_2.

$$S^0(\theta_1,\theta_2)=(w^0)^{-1}\times\begin{cases}\psi_1^0(\theta_1)\,\psi_2^{0T}(\theta_2), & \theta_1<\theta_2\\ \psi_2^0(\theta_1)\,\psi_1^{0T}(\theta_2), & \theta_1>\theta_2\ .\end{cases} \tag{A4}$$

We note finally that the solutions ψ_1^0, ψ_2^0 in Eq. (A2) differ by constant factors from the $\lambda=0$ limit of the power-series solutions for ψ_1, ψ_2 given in Sec. VI.

APPENDIX B: WKB EXPRESSION FOR $\Delta_{\xi j}(\lambda)$

We derive in this Appendix the WKB asymptotic approximation for $\Delta_{\xi j}(\lambda)$ quoted in Sec. V A. Our starting point is the set of coupled differential equations for the components a, c of ψ,

$$\frac{da}{d\theta}-[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\,a+(j+\tfrac{1}{2})(\cos\theta)^{-1}c=0\,,$$
$$\frac{dc}{d\theta}+[\xi(\sin\theta)^{-1}+\lambda\sin\theta]\,c+(j+\tfrac{1}{2})(\cos\theta)^{-1}a=0\,. \tag{B1}$$

These equations are evidently invariant under the interchange

$$\begin{pmatrix}a\\ \xi\\ \lambda\end{pmatrix}\longleftrightarrow\begin{pmatrix}c\\ -\xi\\ -\lambda\end{pmatrix}, \tag{B2}$$

allowing us to obtain equations satisfied by c by a simple substitution once we have found the corresponding equations satisfied by a. Eliminating c and defining a new variable $x=\cos\theta$, we find the following second-order differential equation satisfied by a ($a'=da/dx$, etc.)

$$a''+Pa'+Qa=0\,,$$
$$P=\frac{1-2x^2}{x(1-x^2)}\,,$$
$$Q=\frac{-2\xi\lambda}{1-x^2}-\left[\frac{(j+\frac{1}{2})^2}{x^2(1-x^2)}+\frac{\xi^2}{(1-x^2)^2}+\lambda^2\right]+\frac{\xi}{x(1-x^2)^2}+\frac{\lambda(1-2x^2)}{x(1-x^2)}\,. \tag{B3}$$

Noting that P is unchanged by the substitution of Eq. (B2), we introduce new dependent variables b and d by writing

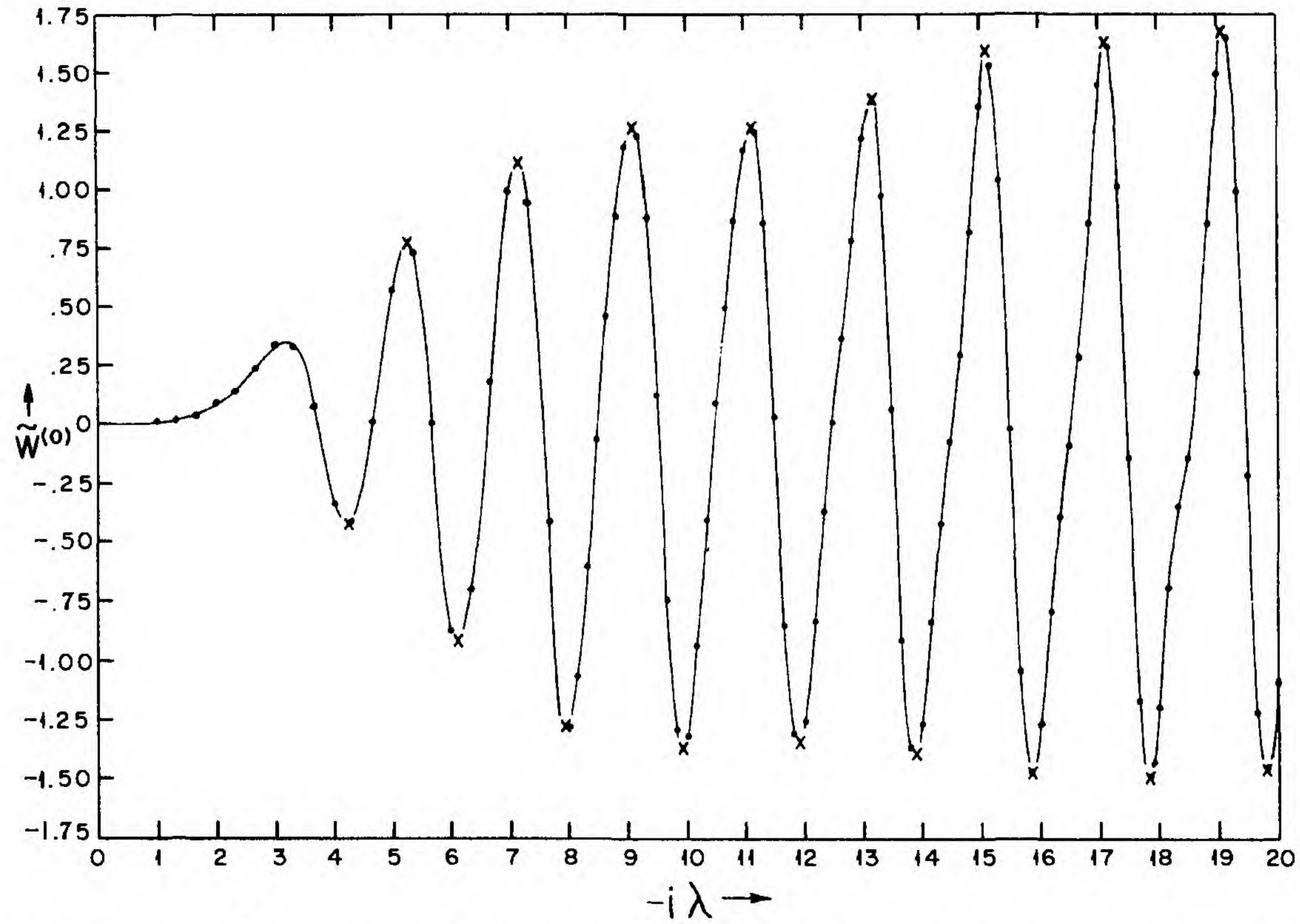


FIG. 6. Results for $\tilde{W}^{(0)}$ versus $-i\lambda$. The dots denote computed points. Maximum and minimum points denoted by × were determined by a polynomial interpolation procedure from the neighboring computed points.

$$\begin{aligned} a &= b\exp\left[-\tfrac{1}{2}\int^x P\,du\right] \\ &= bx^{-1/2}(1-x^2)^{-1/4}, \\ c &= d\exp\left[-\tfrac{1}{2}\int^x P\,du\right] \\ &= dx^{-1/2}(1-x^2)^{-1/4}. \end{aligned} \tag{B4}$$

These satisfy the differential equations

$$\begin{aligned} &b'' + k_b^2 b = 0, \quad d'' + k_d^2 d = 0, \\ &k_b^2 = t_1 + t_2, \quad k_d^2 = t_1 - t_2, \\ &t_1 = \frac{1+2x^2}{4x^2(1-x^2)^2} - \frac{2\xi\lambda}{1-x^2} \\ &\qquad - \left[\frac{(j+\frac{1}{2})^2}{x^2(1-x^2)} + \frac{\xi^2}{(1-x^2)^2} + \lambda^2\right], \\ &t_2 = \frac{\xi}{x(1-x^2)^2} + \frac{\lambda(1-2x^2)}{x(1-x^2)}. \end{aligned} \tag{B5}$$

It is also useful to have the first-order differential equations coupling b and d, which from Eqs. (B1) and (B4) are found to be

$$\begin{aligned} b' + b\left[\frac{1}{2}\,\frac{2x^2-1}{x(1-x^2)} + \frac{\xi}{1-x^2} + \lambda\right] - d\,\frac{(j+\frac{1}{2})}{x(1-x^2)^{1/2}} &= 0, \\ d' + d\left[\frac{1}{2}\,\frac{2x^2-1}{x(1-x^2)} - \frac{\xi}{1-x^2} - \lambda\right] - b\,\frac{(j+\frac{1}{2})}{x(1-x^2)^{1/2}} &= 0. \end{aligned} \tag{B6}$$

Finally, in terms of b and d the Wronskian is given by

$$\begin{aligned} w &= a_2c_1 - a_1c_2 \\ &= \frac{1}{x(1-x^2)^{1/2}}(b_2d_1 - b_1d_2). \end{aligned} \tag{B7}$$

We now proceed to construct approximate, WKB solutions to the above equations when $|\lambda|$ is treated as a large parameter. We begin with the equation for b. We have

$$k_b^2 = -\lambda^2 + \lambda\left[\frac{-2\xi}{1-x^2} + \frac{1-2x^2}{x(1-x^2)}\right] + O(1), \tag{B8}$$

and hence

$$R \equiv \left|\frac{dk_b/dx}{k_b^2}\right| \ll 1 \tag{B9}$$

for all x except very near the end points at $x=0,1$. Near the end points we find

$$\begin{aligned} R &\sim \frac{1}{2x^2|\lambda|^2}, \quad x \approx 0 \\ R &\sim \frac{2\xi+1}{4(1-x)^2|\lambda|^2}, \quad x \approx 1 \end{aligned} \tag{B10}$$

and so except in the intervals

$$x \approx \frac{1}{|\lambda|}, \quad 1-x \approx \frac{\xi^{1/2}}{|\lambda|} \tag{B11}$$

we can use a WKB solution for b and d. Applying the standard lowest-order WKB recipe[12] to the second-order differential equations for b and d, and then imposing the linear equations in Eqs. (B6) which relate b and d, we find the WKB-region solutions

$$\begin{aligned} b_{\rm WKB} &= \frac{A}{\lambda} e^{\lambda x} x^{-1/2}(1+x)^{(\xi-1/2)/2}(1-x)^{-(\xi+1/2)/2} + Be^{-\lambda x} x^{1/2}(1+x)^{-(\xi-1/2)/2}(1-x)^{(\xi+1/2)/2}, \\ d_{\rm WKB} &= \frac{-(j+\frac{1}{2})B}{2\lambda} e^{-\lambda x} x^{-1/2}(1+x)^{-(\xi+1/2)/2}(1-x)^{(\xi-1/2)/2} + \frac{2A}{j+\frac{1}{2}} e^{\lambda x} x^{1/2}(1+x)^{(\xi+1/2)/2}(1-x)^{-(\xi-1/2)/2}. \end{aligned} \tag{B12}$$

In the end-point regions $x \sim 0, 1$ we must join Eq. (B12) on to more accurate approximate solutions. In the vicinity of the end points we find

$$\begin{aligned} k_b^2 &= \frac{-j(j+1)}{x^2} + \frac{\lambda+\xi}{x} - (\lambda+\xi)^2 + \tfrac{3}{4} - j(j+1) + H_0 x + O(x^2), \quad H_0 = -\lambda + 2\xi, \quad x=0 \\ k_b^2 &= \frac{-\frac{1}{4}(\xi-\frac{3}{2})(\xi+\frac{1}{2})}{(1-x)^2} - \frac{\xi\lambda + \frac{1}{2}\lambda - \frac{1}{4}\xi + \frac{1}{4}(\xi-\frac{3}{2})(\xi+\frac{1}{2}) + \frac{1}{2}j(j+1)}{1-x} \\ &\quad - \tfrac{5}{4}[j(j+1) - \tfrac{3}{4}] - [\lambda + \tfrac{1}{4}(\xi-\tfrac{5}{2})]^2 - \tfrac{1}{8}(\xi-\tfrac{3}{2})^2 - \tfrac{1}{8} + H_1(1-x) + O((1-x)^2), \\ &\qquad H_1 = -\tfrac{17}{8}j(j+1) + \xi - \tfrac{1}{8}(\xi+\tfrac{3}{2})(\xi-\tfrac{1}{2}) + \tfrac{9}{8}\lambda - \tfrac{1}{4}\xi\lambda, \quad x=1. \end{aligned} \tag{B13}$$

For $x \approx j/|\lambda|$, the x^{-2}, x^{-1}, x^0 terms near $x=0$ are of order $|\lambda|^2$, whereas the term H_0x is of order $j\xi$, down by a factor $(j/|\lambda|)(\xi/|\lambda|)$ from the leading terms. Similarly, for $1-x \approx \xi/|\lambda|$, the $(1-x)^{-2}$, $(1-x)^{-1}$, $(1-x)^0$ terms near $x=1$ are of order $|\lambda|^2$, with the term $H_1(1-x)$ of order ξ^2, down by a factor of $(\xi/|\lambda|)^2$ from the leading terms. The terms $O(x^2)$ and $O((1-x)^2)$ can be shown to be as small as the linear terms which we have just evaluated. Hence we identify

$$\epsilon_1 = \xi/|\lambda|, \quad \epsilon_2 = j/|\lambda| \tag{B14}$$

as the effective smallness parameters in the WKB solution, and proceed to solve the differential equations at the endpoints neglecting the linear and higher terms in x and $1-x$ in Eqs. (B13). Both at $x=0$ and $x=1$, the differential equations can then be reduced to Whittaker's equation

$$\frac{d^2b}{dz^2}+\left(-\tfrac{1}{4}+\frac{\kappa}{z}+\frac{\tfrac{1}{4}-\mu^2}{z^2}\right)b=0, \tag{B15}$$

with the regular solution

$$b=e^{-z/2}z^{1/2+\mu}\Phi(\tfrac{1}{2}+\mu-\kappa,\,1+2\mu;\,z), \tag{B16}$$

where Φ is the confluent hypergeometric function

$$\Phi(a,c;z)=1+\frac{a}{c}\frac{z}{1!}+\frac{a(a+1)}{c(c+1)}\frac{z^2}{2!}+\cdots$$
$$=e^z\Phi(c-a,c;-z). \tag{B17}$$

Carrying out the solutions explicitly, we find to the required accuracy the following end-point solutions:

$\theta\approx 0,\ x\approx 1$:

$$b_1(z)=\exp\{-[\lambda+\tfrac{1}{4}(\xi-\tfrac{5}{2})]z\}z^{(\xi+1/2)/2}\Phi\left(\xi+\tfrac{1}{2}+\frac{1}{4\lambda}(j+\tfrac{1}{2})^2,\ \xi+\tfrac{1}{2};\ 2[\lambda+\tfrac{1}{4}(\xi-\tfrac{5}{2})]z\right),$$

$$d_1(z)=\frac{-(j+\frac{1}{2})}{2^{1/2}(\xi+\frac{1}{2})}\exp\{-[\lambda+\tfrac{1}{4}(\xi+\tfrac{5}{2})]z\}z^{(\xi+3/2)/2}\Phi(\xi+\tfrac{1}{2},\ \xi+\tfrac{3}{2};\ 2[\lambda+\tfrac{1}{4}(\xi+\tfrac{5}{2})]z),\quad z=1-x;$$

$\theta\approx\frac{1}{2}\pi,\ x\approx 0$: (B18)

$$b_2(x)=e^{-(\lambda+\xi)x}x^{j+1}\Phi(j+\tfrac{1}{2},\ 2(j+1);\ 2(\lambda+\xi)x),$$

$$d_2(x)=e^{-(\lambda+\xi)x}x^{j+1}\Phi(j+\tfrac{3}{2},\ 2(j+1);\ 2(\lambda+\xi)x).$$

Joining the WKB-region solution onto the asymptotic form[13] of the $x\approx 0$ end-point solution, we determine the constants A, B in Eq. (B12) to be

$$\frac{2A}{j+\frac{1}{2}}=\frac{\Gamma(2(j+1))}{\Gamma(j+\frac{3}{2})}(2\lambda)^{-(j+1/2)},$$
$$B=\frac{\Gamma(2(j+1))}{\Gamma(j+\frac{3}{2})}\left(\frac{-1}{2\lambda}\right)^{j+1/2}. \tag{B19}$$

This permits us to extend the solution ψ_2 to the region near $\theta\approx 0,\ x\approx 1$, which is the asymptotic region for the $x\approx 1$ end-point solution ψ_1. Substituting the WKB extension of ψ_2 and the asymptotic expansion of ψ_1 into Eq. (B7), we get for the Wronskian

$$w(\lambda)=-\frac{\Gamma(2(j+1))}{\Gamma(j+\frac{3}{2})}2^{(\xi-1/2)/2}(2\lambda)^{-(j+1/2)}$$
$$\times[e^{\lambda}+e^{-\lambda}(-1)^{j+\xi+1}2^{-2(\xi+1/2)}$$
$$\times(j+\tfrac{1}{2})\Gamma(\xi+\tfrac{1}{2})\lambda^{-(\xi+1/2)}]. \tag{B20}$$

To complete the calculation, we must determine the value $w(0)$ corresponding to the normalization of the solutions ψ_1, ψ_2 used in the above analysis. This is most easily done by a comparison with the explicit free solutions given in Appendix A. Writing

$$\begin{pmatrix}a_1\\c_1\end{pmatrix}_{\lambda=0}=K_1\begin{pmatrix}a_1^0\\c_1^0\end{pmatrix},$$
$$\begin{pmatrix}a_2\\c_2\end{pmatrix}_{\lambda=0}=K_2\begin{pmatrix}a_2^0\\c_2^0\end{pmatrix}, \tag{B21}$$

and letting $\theta\to 0,\ \frac{1}{2}\pi$ to determine K_1, K_2, respectively, we find from Eqs. (B4) and (B18) that

$$K_1=\frac{2^{(\xi-1/2)/2}\Gamma(\xi+\frac{1}{2})\Gamma(j+\frac{3}{2})}{\Gamma(j+\xi+1)},$$
$$K_2=\frac{2^{j+1/2}\Gamma(j+1)\Gamma(\xi+1)}{\Gamma(j+\xi+1)}. \tag{B22}$$

Combining with Eq. (A3) we then get

$$w(0)=-K_1K_2\frac{\Gamma(j+\xi+1)}{\Gamma(\frac{1}{2})\Gamma(\xi+1)\Gamma(j+\frac{3}{2})}. \tag{B23}$$

Dividing Eq. (B20) by Eq. (B23) to get $w(\lambda)/w(0)$, and then using Eq. (4.17), gives the final WKB formula quoted in Eq. (5.2) of the text.

[1] S. L. Adler, Phys. Rev. D 6, 3445 (1972); 7, 3821(E) (1973).
[2] S. L. Adler, Phys. Rev. D 8, 2400 (1973).
[3] We can omit the matrix τ_2 in Eq. (2.32) because the spinors which appear have already been reduced to four-component form.
[4] There is, of course, a third regular singular point at $u=\infty$. For a discussion of the Riemann equation and its solution see G. Birkhoff and G. C. Rota, *Ordinary Differential Equations* (Blaisdell-Ginn, Waltham, Mass., 1969), p. 272 ff.
[5] These may be derived from the identities on pp. 274–276 of Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. 1.

[6]See, for example, G. Birkhoff and G. C. Rota, *Ordinary Differential Equations* (Ref. 4), p. 47.

[7]It is always possible to find solutions satisfying the standardization conditions because, as stressed in Sec. II D, the boundary conditions at $\theta = 0, \frac{1}{2}\pi$ are λ-independent.

[8]Let tr denote the Pauli matrix trace; then Tr denotes the complete trace $\mathrm{Tr}A = \int_0^{\pi/2} d\theta \langle\theta|\mathrm{tr}A|\theta\rangle$.

[9]A. S. B. Holland, *Introduction to the Theory of Entire Functions* (Academic, New York, 1973). See especially Sec. 1.4, Chap. 4, and Sec. 6.2.

[10]S. Coleman (unpublished) has conjectured this to be the case. Coleman argues that at the 45° sector boundaries in Fig. 4, $\mathrm{Re}(\lambda^2)$ changes sign from positive to negative, corresponding to a transition from "magnetic-field-like" to "electric-field-like" behavior of the external-field problem, and suggesting very different analyticity properties on the two sides of the boundary.

[11]A. S. Wightman (unpublished) has proved, in the Minkowski metric case, that the Fredholm determinant can have no zeros for arbitrary purely real external fields.

[12]See, for example L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), third edition, pp. 270–271.

[13]*Higher Transcendental Functions* (Bateman Manuscript Project), edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. 1, p. 278.

Erratum: Massless electrodynamics in the one-photon-mode approximation [Phys. Rev. D 10, 2399 (1974)]

Stephen L. Adler

In Sec. VI B of this paper, numerical evidence was given suggesting that the zeros of $\Delta_{\xi j}(\lambda)$ obey the condition $|\mathrm{Im}\,\lambda|/|\mathrm{Re}\,\lambda|>1$, which would imply cut-plane analyticity for the radiative-corrected vacuum amplitude W_1. Recently, Chernin and Wu[1] have shown that this conjecture is false by deriving the following approximate large-ξ expression for the zeros ξ_n of $\Delta_{\xi\,1/2}(\lambda)$:

$$2[F(\theta_0)+n\pi i] - \tfrac{1}{2}\ln\left(-\frac{2\lambda}{\pi}\right) - \tfrac{3}{2}\ln\cos\theta_0 = 0\,, \tag{1}$$

$$\theta_0 = \sin^{-1}\left(\frac{\xi}{-\lambda}\right)^{1/2}, \quad F(\theta) = \xi\ln\tan\tfrac{1}{2}\theta - \lambda\cos\theta\,.$$

For $\xi = \frac{15}{2}$ this formula gives the following predicted zeros:

$$\begin{aligned} n=2{:}&\quad \lambda = -11.63 + i\,9.58\,,\\ n=3{:}&\quad \lambda = -12.52 + i\,13.14\,,\\ n=4{:}&\quad \lambda = -13.23 + i\,16.57\,. \end{aligned} \tag{2}$$

The $n=4$ zero is the one given in Table II of Sec. VI B; a reexamination of the computer output which I used in preparing Table II indicates that the lower zeros were missed by careless reading of the output (the programming itself was correct), and are indeed given quite accurately by the Chernin-Wu formula. For example, the program used to get Table II gives $\lambda = -11.66 + i\,9.56$ for the location of the $n=2$ zero for $j=\frac{1}{2}$, $\xi = \frac{15}{2}$. For fixed n, the Chernin-Wu formula shows that $-\lambda/\xi \to 1$ as $\xi \to \infty$, and so in fact there are zeros with arbitrarily small $|\mathrm{Im}\,\lambda|/|\mathrm{Re}\,\lambda|$, and hence no zero-free angular sectors for the Fredholm determinant, which is proportional to

$$\prod_{j,\,\xi}[\Delta_{\xi j}(\lambda)\Delta_{\xi j}(-\lambda)]^{2j+1}\,.$$

The zeros "near" the real axis still lie outside the zero-free strip containing the real axis which was established in Sec. IV.

[1]D. Chernin and T. T. Wu (unpublished).

PHYSICAL REVIEW D VOLUME 5, NUMBER 3 1 FEBRUARY 1972

Three-Pion States in the $K_L \to \mu^+\mu^-$ Puzzle

Stephen L. Adler and Glennys R. Farrar*
Institute for Advanced Study, Princeton, New Jersey 08540

and

S. B. Treiman†
National Accelerator Laboratory, Batavia, Illinois 60510
(Received 26 October 1971)

Contributions to the absorptive $K_L \to 2\mu$ amplitude coming from intermediate 3π states are estimated on the basis of recent soft-pion results for the process $3\pi \to 2\gamma$. These contributions turn out to be far too small, by 4 orders of magnitude, to resolve the $K_L \to 2\mu$ puzzle.

All theoretical resolutions so far proposed for the $K_L \to 2\mu$ puzzle[1] are forced to call upon cancellation effects which have to be regarded as accidental at the present level of understanding. Apart from this, the various schemes differ widely with respect to introduction of qualitatively new physics.[2] The most conservative approach is one which dismisses the possibility that CP violation or new kinds of particles or interactions play an important role in the puzzle. Instead, the burden is placed on 3π intermediate states, which are supposed to provide terms which largely cancel the contribution from the 2γ state in the unitarity equation for the absorptive $K_L \to 2\mu$ amplitude. The strain on credulity here lies in the magnitude required of the 3π contribution, a magnitude which has to be appreciably larger than first rough estimates would suggest.[3] In the present note we add our contribution to this strain, in the form of an estimate of 3π contributions based on soft-pion considerations.

In order to assess the 3π effects in a framework which ignores CP violation and accepts standard photon-lepton electrodynamics, one requires information on the amplitudes for $3\pi \to 2\gamma$ and $3\pi \to 2\mu$. In our conventional framework, the latter is fully specified if the former is known for virtual as well as real photons. All the remaining ingredients of a unitarity analysis based on 2γ and 3π intermediate states are well enough known: the $2\gamma \to 2\mu$ amplitude from standard electrodynamic theory, the $K_L \to 2\gamma$ and $K_L \to 3\pi$ amplitudes (or rather, their moduli) from experiment. Throughout the unitarity discussion we ignore all other intermediate states. To lowest order in the fine-structure constant the $2\pi\gamma$ and, strictly speaking, also the $3\pi\gamma$ intermediate states ought to be considered. However, the former has been shown to be unimportant,[4] and the latter can reasonably be expected, on phase-space considerations alone, to be even more negligible. We shall have a brief comment on this later on.

At theoretical issue then are the amplitudes for $3\pi^0$, $\pi^+\pi^-\pi^0 \to$ two real or virtual photons. These objects are of course interesting in their own right, even apart from their role in the $K_L \to 2\mu$ puzzle. In particular, the application of soft-pion considerations has been discussed by Aviv, Hari Dass, and Sawyer[5]; and the subject has since been taken up by other authors.[6-10] Interesting issues concerning current algebra, partial conservation of axial-vector current (PCAC), and Ward-identity anomalies arise here. Especially relevant for our present purposes is the idea, proposed by Aviv and Sawyer,[11] that the soft-pion approximation might provide a reasonable basis for estimating contributions from the 3π states in the unitarity analysis of $K_L \to 2\mu$ decay. It must be said at once that, kinematically, the pions in $K_L \to 3\pi$ decay cannot all three be so very soft, unless one regards the K-meson mass to be "small" on a hadronic scale. With appropriate reservations on this score, one may nevertheless hope that the soft-pion methods provide more reliable estimates than can be gained from purely dimensional and phase-space arguments.

The Aviv-Sawyer analysis[11] of $K_L \to 2\mu$ decay was based on the $3\pi \to 2\gamma$ amplitudes of Refs. 5 and 6. We believe that these amplitude results are in error and that the correct soft-pion expressions are as in Ref. 8. We have therefore repeated the analysis. Despite these corrections, we find with Aviv and Sawyer that the 3π states play a negligible role in the absorptive amplitude for $K_L \to 2\mu$ decay. The "naive" unitarity bound, based solely on the 2γ intermediate state, is corrected at most (depending on phases) by a factor of order 10^{-4} in the decay rate. A brief account follows.

The $K_L \to 2\mu$ amplitude has the structure

$$\mathrm{Amp}(K_L \to 2\mu) = g\bar{u}(p)\gamma_5 v(\bar{p}), \tag{1}$$

where p and $\bar{p}$ denote the μ^- and μ^+ momenta. The

decay rate is given by

$$\Gamma(K_L \to 2\mu) = \frac{M}{8\pi} v |g|^2, \qquad (2)$$

where

$$v = (1 - 4m^2/M^2)^{1/2}$$

is the muon velocity, with m the μ mass, and M the K mass. The object is to estimate the absorptive amplitude Img, on the basis of unitarity considerations, in order to set a lower bound for $K_L \to 2\mu$ decay. To get at the unitarity contribution from the intermediate 2γ state we have to consider $K_L \to 2\gamma$ decay, whose amplitude has the structure

$$\mathrm{Amp}(K_L \to 2\gamma) = G \epsilon_{\mu\nu\rho\sigma} \epsilon_\mu^{(1)} k_\nu^{(1)} \epsilon_\rho^{(2)} k_\sigma^{(2)}, \qquad (3)$$

where $k^{(i)}$ and $\epsilon^{(i)}$ are the momentum and polarization vectors of the ith photon. The decay rate is

$$\Gamma(K_L \to 2\gamma) = \frac{M^3}{64\pi} |G|^2. \qquad (4)$$

The contribution to Img coming from the 2γ state is given by

$$\mathrm{Im}g|_{2\gamma} = \frac{m\alpha}{4v} \ln\left(\frac{1+v}{1-v}\right) \mathrm{Re}G. \qquad (5)$$

Now the modulus $|G|$ is known from empirical information on the $K_L \to 2\gamma$ decay rate. If unitarity contributions coming from 3π states are systematically ignored for both $K_L \to 2\gamma$ and $K_L \to 2\mu$ decay, then $\mathrm{Im}g = \mathrm{Im}g|_{2\gamma}$, $\mathrm{Re}G = |G|$, and one finds the "naive" unitarity bound

$$\Gamma(K_L \to 2\mu)/\Gamma(K_L) \gtrsim 6 \times 10^{-9}. \qquad (6)$$

Our task here is to compute the direct 3π contributions to Img, and also their contributions to ImG. For these purposes we require the amplitudes for $3\pi^0$, $\pi^+\pi^-\pi^0 \to$ two real or virtual photons. We adopt, but do not reproduce here, the soft-pion expressions of Ref. 8. These expressions contain three parameters, of which two are well established experimentally: F^π, the constant which describes $\pi^0 \to 2\gamma$ decay; and f, the PCAC constant. The remaining parameter, x, measures the isotensor component of the "σ term" in the current-algebra treatment of π-π scattering. One usually supposes, as we shall do here, that $x = 0$. Unless x is unbelievably large, of order 10^3–10^4, this neglect will not qualitatively alter our conclusion that the 3π states do not resolve the $K_L \to 2\mu$ puzzle. Indeed, given the formulas of Ref. 8, and with a little thought about the structure of the unitarity equations and the size of phase space for three pions, one can readily arrive at this qualitative conclusion from rough dimensional arguments. Nevertheless, since we have in fact carried out the numerical work in detail, and because a certain delicacy appears in the details, we shall comment here on a few technical points. For the unitarity calculations we require not only the $3\pi \to 2\gamma$ and $3\pi \to 2\mu$ amplitudes, but also the full complex amplitudes for $K_L \to 3\pi$. The latter are known from experiment only in modulus. However, we can get upper bounds on the 3π contributions by replacing all amplitudes in the unitarity equations with their moduli. It is these upper bounds that we shall report. The computations for ImG are now completely straightforward. For the $3\pi^0$ and $\pi^+\pi^-\pi^0$ contributions we find

$$\begin{aligned} \mathrm{Im}G|_{3\pi^0} &\lesssim 3 \times 10^{-5} |G|, \\ \mathrm{Im}G|_{\pi^+\pi^-\pi^0} &\lesssim 2 \times 10^{-5} |G|. \end{aligned} \qquad (7)$$

It is evident that the 3π effects here are totally negligible.

Computation of the direct 3π contributions to Img, the absorptive $K_L \to 2\mu$ amplitude, is somewhat less straightforward. The formulas of Ref. 8 are supposed to apply (in the soft-pion limit) for virtual as well as real photons, and they therefore provide a basis for computation of the $3\pi \to 2\mu$ amplitude. On inspection of the formulas for $3\pi \to$ two real or virtual photons one observes two kinds of terms: those which describe emission of a photon by an external pion (bremsstrahlung terms) and those which do not. Correlation of these descriptive expressions with explicit terms in the formulas should be evident and is left to the reader. The $3\pi^0 \to 2\gamma$ amplitude is purely of the nonbremsstrahlung type, whereas the $\pi^+\pi^-\pi^0$ amplitude has both kinds of terms. Computation of the bremsstrahlung-term contributions to $3\pi \to 2\mu$ presents no difficulties, although it is tedious. The calculation here has a structure of the kind associated with a one-loop box diagram and was carried out numerically. For practical purposes we found it convenient to use dispersion-relation methods, taking the invariant squared mass of the 3π system as the dispersion variable. One encounters no anomalous thresholds here, thanks to the masslessness of the physical photons in the intermediate state $3\pi \to 2\gamma \to 2\mu$. For the nonbremsstrahlung terms, the calculation of the $3\pi \to 2\mu$ amplitude has a structure of the kind associated with a one-loop triangle diagram. But here one encounters a logarithmically divergent integral. This comes about because the corresponding amplitudes for $3\pi \to$ two virtual photons do not have any damping as the virtual-photon masses become very large. The soft-pion approximation is unsatisfactory in this regard. However, since the divergence is only logarithmic, we do not think it misleading to employ a cutoff. We again employ dispersion-relation methods. The dispersion integral is loga-

rithmically divergent and we simply cut it off, at an invariant squared mass taken rather arbitrarily to be 1 GeV^2.

Once the $3\pi \to 2\mu$ amplitudes have been estimated, computation of the 3π contributions to the absorptive $K_L \to 2\mu$ amplitude is now a simple matter. We present the results in the form of comparison of the 3π and 2γ contributions to $\mathrm{Im}g$,

$$\begin{aligned} \mathrm{Im}g|_{3\pi^0} &\lesssim 3\times 10^{-5}\,\mathrm{Im}g|_{2\gamma}, \\ \mathrm{Im}g|_{\pi^+\pi^-\pi^0} &\lesssim 3\times 10^{-5}\,\mathrm{Im}g|_{2\gamma}. \end{aligned} \tag{8}$$

In summary, the 3π states, at least when treated in the soft-pion approximation, do nothing to resolve the $K_L \to 2\mu$ puzzle.[12]

Finally, we comment briefly on the $\pi^+\pi^-\pi^0\gamma$ intermediate state, which is the remaining intermediate state which can contribute at this order in α. Although the decay $K_L \to \pi^+\pi^-\pi^0\gamma$ has not been observed, to leading order in the photon momentum (the bremsstrahlung approximation) the amplitude for this process is related by gauge invariance to the amplitude for $K_L \to \pi^+\pi^-\pi^0$. The current-algebra coupling[6] of a photon to $\pi^+\pi^-\pi^0$ can then be used to compute $\pi^+\pi^-\pi^0\gamma \to \mu^+\mu^-$. An estimate of the relevant integrations indicates a contribution to $\mathrm{Im}g$ of essentially the same size as that coming from the 3π intermediate state. So the $3\pi\gamma$ contribution is also at least four orders of magnitude too small to resolve the $K_L \to 2\mu$ puzzle.

We wish to thank Professor R. F. Dashen for a helpful discussion. One of the authors (SLA) also wishes to acknowledge a pleasant visit at the National Accelerator Laboratory, where part of this work was done.

---

*National Science Foundation Postdoctoral Fellow.
†Permanent address: Physics Department, Princeton University, Princeton, N. J. 08540.

[1]A. R. Clark, T. Elioff, R. C. Field, H. J. Frisch, R. P. Johnson, L. T. Kerth, and W. A. Wenzel, Phys. Rev. Letters 26, 1667 (1971).
[2]G. R. Farrar and S. B. Treiman, Phys. Rev. D 4, 257 (1971); N. Christ and T. D. Lee, *ibid.* 4, 209 (1971); M. K. Gaillard, Phys. Letters 35B, 431 (1971); 36B, 114 (1971); H. H. Chen and S. Y. Lee, Phys. Rev. D 4, 903 (1971); B. R. Martin, E. de Rafael, J. Smith, and Z. E. S. Uy, *ibid.* 4, 913 (1971); L. Wolfenstein (unpublished); H. H. Chen, K. Kawarabayashi, and G. L. Shaw, Phys. Rev. D 4, 3514 (1971).
[3]B. R. Martin, E. de Rafael, and J. Smith, Phys. Rev. D 2, 179 (1970). In fact, even at the rough dimensional level one would suspect that the bound given in this reference is a substantial overestimate of the actual 3π contribution.
[4]See, under Ref. 2, the papers by Gaillard and by Farrar and Treiman.
[5]R. Aviv, N. D. Hari Dass, and R. F. Sawyer, Phys. Rev. Letters 26, 591 (1971).
[6]R. Aviv and R. F. Sawyer, Phys. Rev. D 4, 451 (1971); Tsu Yao, Phys. Letters 35B, 225 (1971).
[7]E. S. Abers and S. Fels, Phys. Rev. Letters 26, 1512 (1971).
[8]S. L. Adler, B. W. Lee, S. B. Treiman, and A. Zee, Phys. Rev. D 4, 3497 (1971).
[9]T. F. Wong, Phys. Rev. Letters 27, 1617 (1971).
[10]J. Wess and B. Zumino, Phys. Letters 37B, 95 (1971).
[11]R. Aviv and R. F. Sawyer, Phys. Rev. D 4, 2740 (1971).
[12]After this work was completed, we received a report on the same subject from M. Pratap, J. Smith, and Z. E. S. Uy [Phys. Rev. D 5, 269 (1972).] Our conclusions are similar.

PHYSICAL REVIEW D VOLUME 10, NUMBER 11 1 DECEMBER 1974

Some simple vacuum-polarization phenomenology: $e^+e^- \rightarrow$ hadrons; the muonic-atom x-ray discrepancy and $g_\mu - 2$

Stephen L. Adler
National Accelerator Laboratory, Batavia, Illinois 60510
and The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 25 July 1974)

We give a simple phenomenological analysis of hadronic and electronic vacuum-polarization effects. We argue that the derivative of the hadronic vacuum polarization, evaluated in the spacelike region, provides a useful meeting ground for comparing $e^+e^- \rightarrow$ hadron annihilation data (assumed to arise from one-photon annihilation) with the predictions of parton models and of asymptotically free field theories. Using dispersion relations to connect the annihilation and spacelike regions, we discuss the implications in the spacelike region of a constant e^+e^- annihilation cross section. In particular, we show that a flat cross section between $t = 25$ and $t = 81$ $(\mathrm{GeV}/c)^2$ would provide strong evidence against a precociously asymptotic "color" triplet model for hadrons. We then turn to a consideration of the apparent discrepancy between observed and calculated muonic-atom x-ray transition energies. Specifically, we analyze the hypothesis of attributing this discrepancy to a deviation of the asymptotic electronic vacuum polarization from its expected value, a possibility which is compatible with all current high-precision tests of quantum electrodynamics. Under the additional technical assumption that the postulated discrepancy in the electronic vacuum-polarization spectral function increases monotonically with t, the hypothesis predicts a *decrease* in the expected value of the muon-magnetic-moment anomaly $a_\mu = \frac{1}{2}(g_\mu - 2)$ of at least -0.96×10^{-7}, which should be detectable in the next round of $g_\mu - 2$ experiments and which is substantially larger than likely uncertainties in the hadronic contribution to a_μ. By contrast, postulating a weakly coupled scalar boson ϕ to explain the muonic-atom discrepancy would imply a (very small) increase in the expected value of a_μ. Both the vacuum-polarization and scalar-boson hypotheses (for $M_\phi \geq 1$ MeV) predict a reduction of order 0.027 eV in the $2p_{1/2} - 2s_{1/2}$ transition energy in $[^4\mathrm{He}, \mu]^+$, an effect which may be observable.

I. INTRODUCTION

A number of recent experiments have brought aspects of vacuum-polarization phenomena to the fore. Most prominent are the measurements by the Cambridge Electron Accelerator (CEA) and the Stanford Linear Accelerator Center–Lawrence Berkeley Laboratory (SLAC–LBL) groups of an unexpectedly large cross section for $e^+e^- \rightarrow$ hadrons,[1] which gives the absorptive part of the hadronic vacuum polarization. In another area of physics, measurements of muonic-atom x-ray transition energies, undertaken to probe the asymptotic form of the electronic vacuum polarization, appear to show a persistent deviation from theoretical expectations.[2] Forthcoming high-precision measurements of the muon-magnetic-moment anomaly $g_\mu - 2$ will provide an even more sensitive probe of the asymptotic electronic vacuum polarization, and of the hadronic vacuum polarization as well. We present in this paper simple phenomenological arguments which bear on the interpretation of both the annihilation and the muonic experiments. Although fundamentally different physical issues are at stake in the two classes of experiments, common elements of formalism make it natural to consider them together. In Sec. II we use dispersion relations to determine what the timelike-region e^+e^- annihilation data say about the possibility of precocious asymptotic scaling in the spacelike region of the hadronic vacuum polarization (assuming that the observed data do indeed result from one-photon annihilation). In Sec. III we analyze the muonic experiments, with the aim of distinguishing between the possibilities that the muonic-atom x-ray discrepancies may arise from a discrepancy in the asymptotic electronic vacuum polarization, or from the existence of a weakly coupled light scalar boson. Some technical details are given in the appendixes.

II. ELECTRON-POSITRON ANNIHILATION AND PRECOCIOUS SPACELIKE SCALING

The experimental data for electron-positron annihilation into hadrons are conveniently expressed in terms of the ratio $R(t)$, defined as

$$R(t) = \frac{\sigma(e^+e^- \rightarrow \text{hadrons}; t)}{\sigma(e^+e^- \rightarrow \mu^+\mu^-; t)}, \tag{1}$$

with

$$\sigma(e^+e^- \to \mu^+\mu^-; t) = \left(1+\frac{2m_\mu^2}{t}\right)\left(1-\frac{4m_\mu^2}{t}\right)^{1/2}\frac{4\pi\alpha^2}{3t}$$

$$\approx \frac{4\pi\alpha^2}{3t} = \frac{87\times10^{-33}\ \mathrm{cm}^2}{t\ [\mathrm{in}\ (\mathrm{GeV}/c)^2]} \tag{2}$$

and with t the virtual-photon four-momentum squared. In Fig. 1 we have plotted (versus $E = t^{1/2}$) a smooth interpolation through all available experimental data for R in the continuum region (excluding the ρ, ω, and ϕ vector-meson contributions). The CEA and SLAC–LBL data points are indicated,[1] while the portion of the curve below $t = 2.5$ is taken from the "eyeball" fit given by Silvestrini.[3] When replotted versus t, the data for $R(t)$ rise approximately linearly, indicating a roughly constant hadronic annihilation cross section of 21×10^{-33} cm^2. Assuming that single-photon annihilation is indeed being measured, this behavior strongly contradicts the asymptotic behavior expected on the basis of parton or of asymptotically free-field-theory models of the hadrons, which predict

$$R \sim C, \quad t\to\infty \tag{3}$$

with the constants C tabulated in Table I. However, it can always be argued that while precocious asymptotic behavior is expected from the SLAC scaling results in the *spacelike* region, the annihilation reaction involves the timelike region, in which asymptotic predictions may be approached much more slowly. This objection naturally raises the question of determining what the annihilation data tell us about behavior in the spacelike region.

To answer this question we consider the renormalized hadronic vacuum-polarization tensor ($t = q^2$)

$$\Pi_{\mu\nu}^{(H)}(q) = (q_\mu q_\nu - t g_{\mu\nu})\Pi^{(H)}(t), \tag{4}$$

which obeys the dispersion relation

$$\Pi^{(H)}(t) = t\int_{4m_\pi^2}^\infty \frac{du}{u}\,\frac{(1/\pi)\,\mathrm{Im}\Pi^{(H)}(u)}{u-t}, \tag{5}$$

and which is related to the electron-positron annihilation cross section into hadrons by

$$\sigma(e^+e^-\to \mathrm{hadrons}; u) = \frac{1}{u}\mathrm{Im}\Pi^{(H)}(u)\times(\text{known constants}). \tag{6}$$

Rather than using Eq. (5) directly, we consider its first derivative

$$\frac{d}{dt}\Pi^{(H)}(t) = \int_{4m_\pi^2}^\infty du\,\frac{(1/\pi)\,\mathrm{Im}\Pi^{(H)}(u)}{(u-t)^2}, \tag{7}$$

which on substituting Eq. (6) and using Eqs. (1) and (2) can be rewritten as

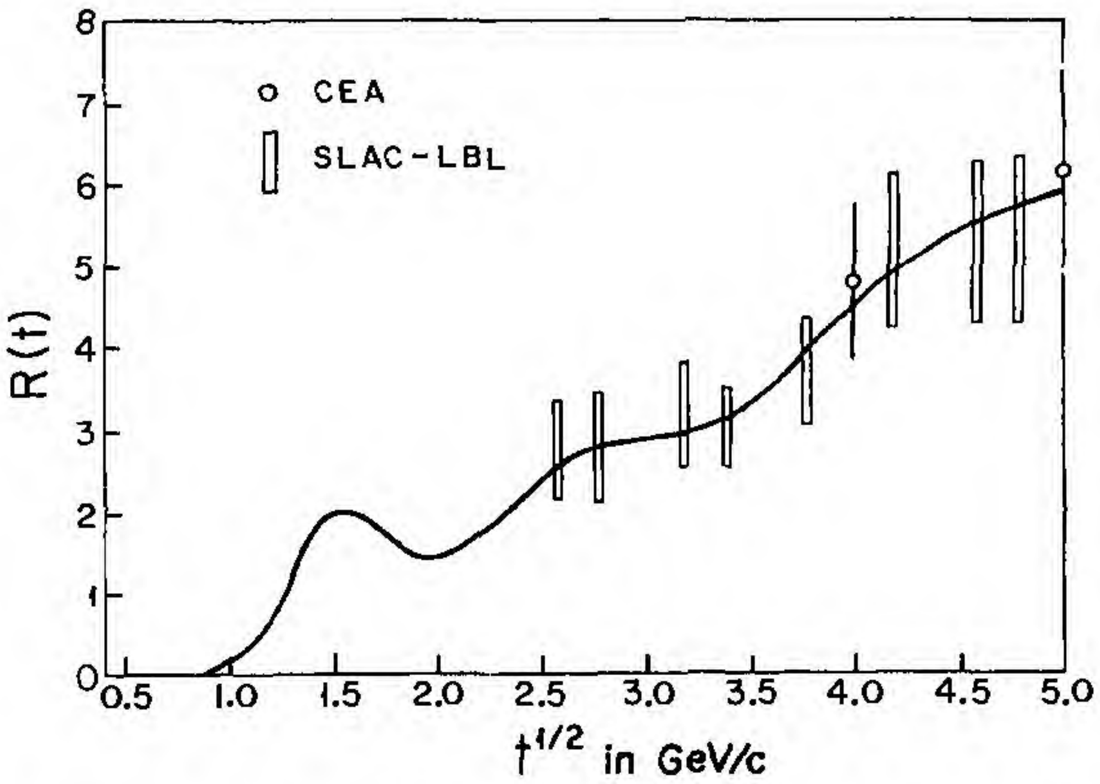


FIG. 1. "Eyeball" fit to the continuum e^+e^- annihilation data. The ρ, ω, and ϕ vector-meson contributions are not included.

$$\frac{d}{dt}\Pi^H(t) = \int_{4m_\pi^2}^\infty du\,\frac{R(u)}{(u-t)^2}\times(\text{known constants}). \tag{8}$$

Restricting ourselves to the spacelike region $t=-s$, $s>0$ and rescaling to remove the constant factors, we obtain from Eq. (8) our basic relation

$$T(-s) \equiv \int_{4m_\pi^2}^\infty \frac{du\,R(u)}{(s+u)^2}$$

$$= \frac{d}{dt}\Pi^{(H)}(t)\Big|_{t=-s}\times(\text{known constants}). \tag{9}$$

The quantity $T(-s)$ has two desirable properties which make it suitable for studying the implications of the annihilation reaction for spacelike-region behavior:

(i) The integrand in Eq. (9) is *positive definite*, and so omitting the high-energy tail of the integral makes an error of known sign. Specifically, if experimental data on R are available only up to a maximum momentum transfer squared t_C, and if we define $T_{\mathrm{obs}}(-s)$ by

$$T_{\mathrm{obs}}(-s) = \int_{4m_\pi^2}^{t_C}\frac{du\,R(u)}{(s+u)^2}, \tag{10}$$

TABLE I. Values of C in different models.

Model	C
Simple quark triplet	$\frac{2}{3}$
Color quark triplet	2
Color quark quartet	$\frac{10}{3}$
Han-Nambu triplet	4
Han-Nambu quartet	6

then we have

$$T_{obs}(-s) \leq T(-s), \quad 0 \leq s < \infty. \tag{11}$$

(ii) It is the quantity $T(-s)$ for which parton models and asymptotically free field theories most directly make predictions[4]; the asymptotic predictions for R are always obtained *from* the prediction for $T(-s)$ by a dispersion-relation argument, which is bypassed if we use $T(-s)$ as the primary phenomenological object. In a model in which R asymptotically approaches C, we have

$$T_{th}(-s) \sim C/s, \quad s \to \infty. \tag{12}$$

In asymptotically free field theories, the leading logarithmic correction to Eq. (12) is also determined. Specifically, in the SU(3)⊗SU(3)′ "color" triplet model of the hadrons, one has[4]

$$T_{th}(-s) \sim \frac{2}{s}\left(1 + \frac{4}{9}\frac{1}{\ln(s/s_0)} + \cdots\right), \tag{13}$$

with s_0 an arbitrary momentum scale which, in the numerical work, we will take as 2 $(\mathrm{GeV}/c)^2$.

Before proceeding to numerical applications, let us briefly discuss the question of subtractions. Clearly, if the one-photon annihilation cross section were to remain constant as $t \to \infty$, we would have $R(u) \propto u$ as $u \to \infty$ and the integral in Eq. (9) would need an additional subtraction to be well defined.[5] However, such behavior of R would in itself contradict Eq. (3) for all values of C, and hence would rule out all versions of the parton model or of asymptotically free field theories. On the other hand, if Eq. (3) is true for any finite C, then the integral in Eq. (9) converges as it stands and provides a suitable medium for comparing the annihilation data with theoretical expectations in the spacelike region. Note that a constant subtraction term in Eq. (5), which would be present if we renormalize at a point other than $t = 0$, would not contribute to the t derivative in Eq. (7); hence the renormalization prescription is not a possible source of ambiguity.

We turn now to the numerical results. In Fig. 2 we plot $T_{obs}(-s)$ [in units where unity = (1 GeV/c)^2], as obtained from all experimental data up to $t_C = 25$ $(\mathrm{GeV}/c)^2$ according to the formula[6]

$$T_{obs}(-s) = T^{\omega+\phi}(-s) + T^{\rho}(-s) + T^{\mathrm{cont}(1)}(-s),$$

$$T^{\omega+\phi}(-s) = \frac{9\pi}{\alpha^2}\sum_{V=\omega,\phi}\frac{M_V\Gamma(V\to e^+e^-)}{(s+M_V^2)^2},$$

$$T^{\rho}(-s) = \int_{4m_\pi^2}^{\infty}\frac{dt}{(s+t)^2}\frac{1}{4}\left(1-\frac{4m_\pi^2}{t}\right)^{3/2}|F_\pi(t)|^2, \tag{14}$$

$$T^{\mathrm{cont}(1)}(-s) = \int_{0.39}^{25}\frac{dt}{(s+t)^2}R(t).$$

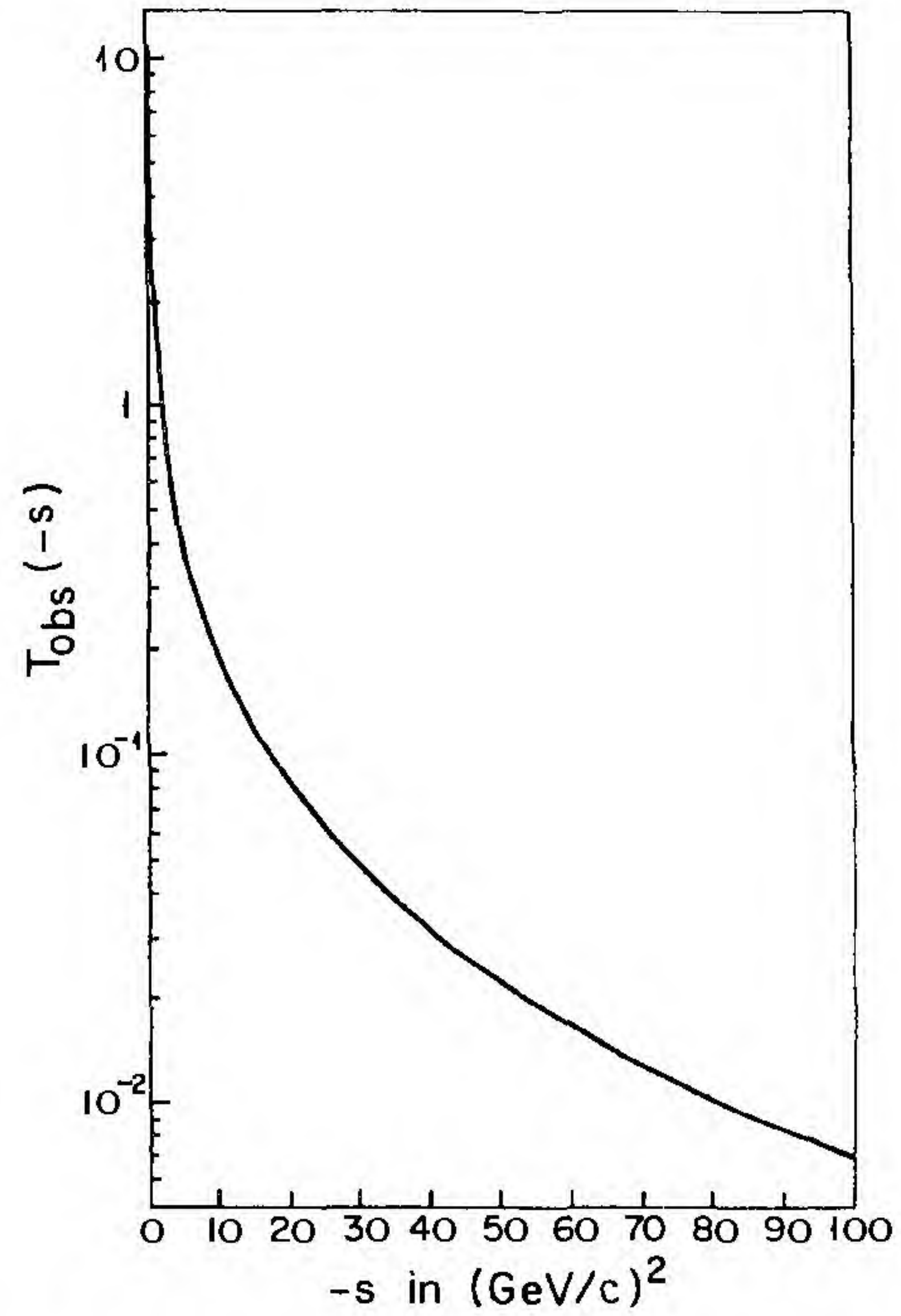


FIG. 2. The function $T_{obs}(-s)$ as obtained from all experimental data up to $t_C = 25$ (Gev/c)^2, in units where unity = (1 GeV/c)^2.

The vector-meson parameters appearing in Eq. (14) are given in Appendix A, while $R(t)$ is the continuum contribution to R graphed in Fig. 1. In Fig. 3 we plot a family of curves, obtained by assuming that for $25 \leq t \leq t_C$ the annihilation cross section $\sigma(e^+e^- \to \mathrm{hadrons}; t)$ remains constant at 21×10^{-33} cm^2. That is, we take

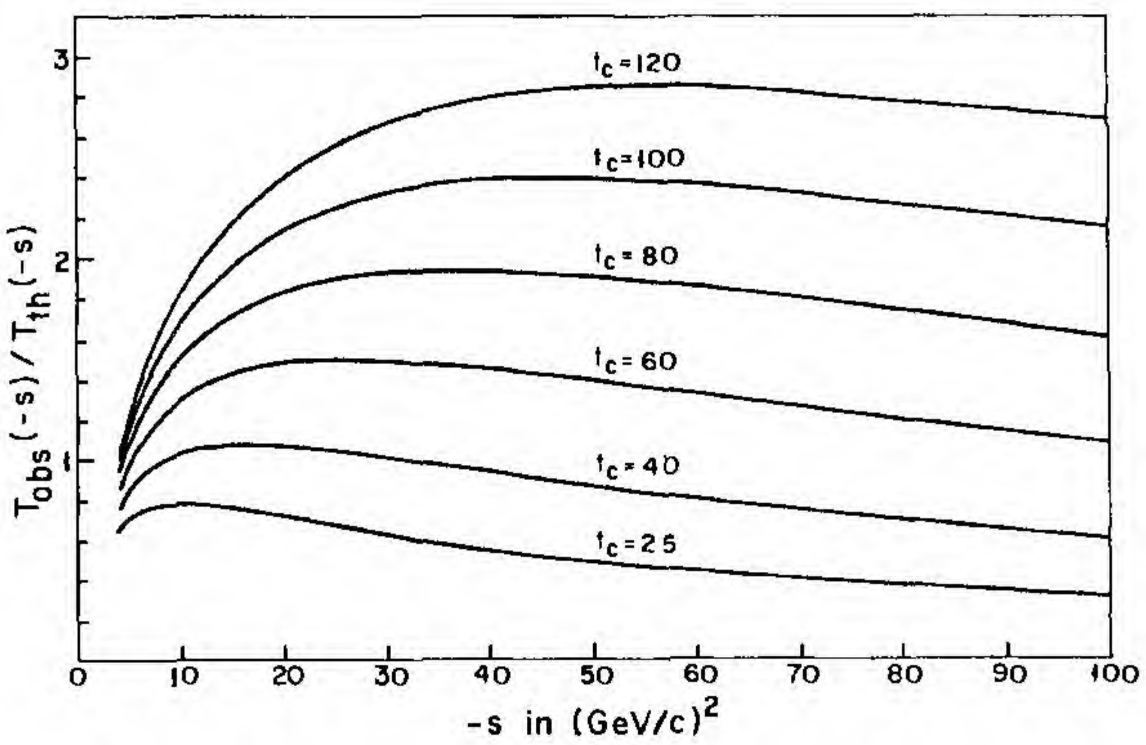


FIG. 3. Ratios of $T_{obs}(-s)$ to $T_{th}(-s)$, with $T_{th}(-s)$ the color-triplet prediction of Eq. (13). The $t_C = 25$ curve uses the presently known data; the curves for higher t_C assume a constant hadronic annihilation cross section of 21×10^{-33} cm^2 above 25 (GeV/c)^2.

$$T_{\rm obs}(-s) = T^{\omega+\phi}(-s) + T^{\rho}(-s) + T^{\rm cont(1)}(-s) + T^{\rm cont(2)}(-s)\,,$$
$$T^{\rm cont(2)}(-s) = \int_{25}^{t_C} \frac{dt}{(s+t)^2} \times 5.94\left(\frac{t}{25}\right) = 0.24\left[\ln\left(\frac{s+t_C}{s+25}\right) - s\frac{(t_C-25)}{(s+t_C)(s+25)}\right]. \tag{15}$$

Rather than plotting $T_{\rm obs}(-s)$ we have plotted the comparison ratio $T_{\rm obs}(-s)/T_{\rm th}(-s)$, with $T_{\rm th}(-s)$ the "color" triplet prediction of Eq. (13). The $t_C=25$ curve is just the curve of Fig. 2 divided by Eq. (13); since this curve lies below 1, the existing annihilation data do not yet challenge the "color" triplet model in the spacelike region. (However, since the $t_C=25$ curve lies well above $\frac{1}{3}$, the existing data already definitively rule out a precociously asymptotic simple quark-triplet model.) Evidently, the curves in Fig. 3 rise rapidly with t_C and show that if the annihilation cross section should remain constant at roughly 21×10^{-33} cm^2 in the region $25 \le t \le 81$, which will be accessible at SPEAR II, a precociously asymptotic "color" triplet model would be ruled out in the spacelike region.

To explore the consequences of an annihilation cross section which remains flat up to large t_C, we ignore the vector-meson contributions to $T_{\rm obs}$ and approximate $T^{\rm cont(1)}(-s)$ by taking $R(t)\approx 0$, $t<2$; $R(t)\approx 0.24t$, $2\le t\le 25$, giving the simple analytic expression

$$T_{\rm obs}(-s) = \int_2^{t_C}\frac{dt}{(s+t)^2}\times 0.24t = 0.24\left[\ln\left(\frac{s+t_C}{s+2}\right) - s\frac{(t_C-2)}{(s+t_C)(s+2)}\right] \approx 0.24\left[\ln\left(\frac{s+t_C}{s}\right) - \frac{t_C}{s+t_C}\right]. \tag{16}$$

Hence,

$$\frac{T_{\rm obs}(-s)}{1/s} \approx R(t_C) f(z)\,,$$
$$R(t_C) = 0.24 t_C\,, \tag{17}$$
$$f(z) = \tfrac{1}{2}\ln(1+z) - \frac{1}{1+z}\,, \quad z = t_C/s\,.$$

A simple maximization shows that $f(z)$ attains a maximum of 0.22 at $z_M^{-1} = s_M/t_C = 0.46$, and falls to half maximum at $z_L^{-1} = s_L/t_C = 0.053$ and $z_U^{-1} = s_U/t_C = 3.22$. That is, $T_{\rm obs}(-s)/s^{-1}$ reaches a maximum value

$$[T_{\rm obs}(-s)/s^{-1}]^{\max} = 0.22\times 0.24 t_C = 0.053 t_C\,, \tag{18a}$$

and lies above half this value in the wide range

$$0.053 t_C \le s \le 3.22 t_C\,. \tag{18b}$$

To give a concrete illustration, if $\sigma(e^+e^- \to \text{hadrons}; t)$ should remain constant up to the maximum t_C of 900 obtainable in a 15 GeV/c on 15 GeV/c storage ring, the maximum of $T_{\rm obs}(-s)/s^{-1}$ would be $0.053\times 900 = 48$. This would exclude by a factor of 2 parton or asymptotically free models with $C\le 24$, thus covering just about every model which has been seriously proposed.

III. MUONIC-ATOM X-RAY DISCREPANCY AND $g_\mu - 2$

Recent studies of the transition energies between large circular orbits in muonic atoms have shown persistent discrepancies between theory and experiment. Because the muonic orbits in question lie well outside the nucleus and well inside the innermost K-shell electrons, one believes that nuclear size and electron screening corrections can be reliably estimated. In particular, the disputed nuclear-size corrections to the vacuum-polarization potential have been reevaluated recently by three independent groups,[7] in good agreement with one another. A survey of all known theoretical corrections has been given by Watson and Sundarasen[2] (see also Rafelski *et al.*[8]), with the conclusion that all important effects within the standard electrodynamic theory have been correctly taken into account. On the experimental side, independent measurements by the groups of Dixit *et al.*[9] and of Walter *et al.*[10] agree on x-ray transition energies which deviate by 2 standard deviations from the theoretical predictions, as summarized in Table II. While it may still turn out that systematic experimental errors or errors or omissions in the theoretical calculations account for the discrepancy, we will assume this not to be the case. Rather, we will treat the discrepancy as a real effect, to be explained by modifications in the conventional theory.

The unique aspect of the muonic-atom transition energies is that, because the muonic orbits lie well inside the electron Compton wavelength, they receive a large contribution from the electronic vacuum-polarization potential and (unlike the more accurate Lamb-shift experiments) they probe the *asymptotic structure* of this potential. Motivated by this observation, our principal focus will be to explore the possibility that the observed x-ray energy discrepancy arises from a nonperturbative deviation of the electronic vacuum polarization from its expected value. Such an effect is qualitatively expected (but with unknown quantitative form) if recent speculations that the fine-

TABLE II. Muonic atom x-ray discrepancies.

Element $_ZE$	Transition $^nl_j \rightarrow ^{n-1}l-1_{j-1}$	$-\delta E_\gamma \equiv E_\gamma(\text{th}) - E_\gamma(\text{expt})$ (eV)	Average discrepancy $-\delta \bar{E}_\gamma$ (eV)	Reduced average discrepancy $-\delta \bar{E}_\gamma = \frac{-\delta E_\gamma}{4.35\ \text{eV} \times Z^2[1/(n-1)^2 - 1/n^2]}$
$_{20}$Ca	$^3d_{3/2} \rightarrow ^2p_{1/2}$ $^3d_{5/2} \rightarrow ^2p_{3/2}$	7 ± 19 11 ± 17	9 ± 13	$(37.2 \pm 54) \times 10^{-3}$
$_{22}$Ti	$^3d_{3/2} \rightarrow ^2p_{1/2}$ $^3d_{5/2} \rightarrow ^2p_{3/2}$	-3 ± 19 10 ± 18	3.5 ± 13	$(12.0 \pm 45) \times 10^{-3}$
$_{26}$Fe	$^3d_{3/2} \rightarrow ^2p_{1/2}$ $^3d_{5/2} \rightarrow ^2p_{3/2}$	21 ± 20 10 ± 17	15.5 ± 13	$(37.9 \pm 32) \times 10^{-3}$
$_{38}$Sr	$^4f_{5/2} \rightarrow ^3d_{3/2}$ $^4f_{7/2} \rightarrow ^3d_{5/2}$	11 ± 20 0 ± 18	5.5 ± 13	$(18.0 \pm 43) \times 10^{-3}$
$_{47}$Ag	$^4f_{5/2} \rightarrow ^3d_{3/2}$ $^4f_{7/2} \rightarrow ^3d_{5/2}$	27 ± 20 19 ± 20	23 ± 14	$(49.3 \pm 30) \times 10^{-3}$
$_{48}$Cd	$^4f_{5/2} \rightarrow ^3d_{3/2}$ $^4f_{7/2} \rightarrow ^3d_{5/2}$	13 ± 19 7 ± 17	10 ± 13	$(20.5 \pm 27) \times 10^{-3}$
$_{50}$Sn	$^4f_{5/2} \rightarrow ^3d_{3/2}$ $^4f_{7/2} \rightarrow ^3d_{5/2}$	21 ± 21 25 ± 19	23 ± 14	$(43.5 \pm 26) \times 10^{-3}$
$_{56}$Ba	$^4f_{5/2} \rightarrow ^3d_{3/2}$ $^4f_{7/2} \rightarrow ^3d_{5/2}$	55 ± 23 76 ± 20	65.5 ± 15	$(98.8 \pm 23) \times 10^{-3}$
	$^5g_{7/2} \rightarrow ^4f_{5/2}$ $^5g_{9/2} \rightarrow ^4f_{7/2}$	12 ± 17 3 ± 16	7.5 ± 12	$(24.4 \pm 39) \times 10^{-3}$
$_{80}$Hg	$^5g_{7/2} \rightarrow ^4f_{5/2}$ $^5g_{9/2} \rightarrow ^4f_{7/2}$	52 ± 24 38 ± 25	45 ± 17	$(71.8 \pm 27) \times 10^{-3}$
$_{81}$Tl	$^5g_{7/2} \rightarrow ^4f_{5/2}$ $^5g_{9/2} \rightarrow ^4f_{7/2}$	31 ± 24 40 ± 24	35.5 ± 17	$(55.3 \pm 26) \times 10^{-3}$
$_{82}$Pb	$^5g_{7/2} \rightarrow ^4f_{5/2}$ $^5g_{9/2} \rightarrow ^4f_{7/2}$	52 ± 21 45 ± 18	48.5 ± 14	$(73.7 \pm 21) \times 10^{-3}$

structure constant α is electrodynamically determined prove to be correct.[11] We will also briefly consider an alternative explanation which has been advanced to explain the x-ray discrepancy, the possible existence of a weakly coupled light scalar boson.[12]

To calculate the effects of a possible discrepancy in the electronic vacuum polarization we start from the Uehling potential written in spectral form,

$$V(r) = -Z\frac{\alpha^2}{3\pi}\int_{4m_e^2}^{\infty}\frac{dt}{t}\left(\frac{e^{-t^{1/2}r}}{r}\right)\rho_e[t],$$

$$\rho_e[t] = \left(1 + \frac{2m_e^2}{t}\right)\left(1 - \frac{4m_e^2}{t}\right)^{1/2}. \quad (19)$$

If we now assume that the spectral function $\rho_e[t]$ is changed by nonperturbative effects[13] to $\rho_e[t] + \delta\rho[t]$, then V is replaced by $V + \delta V$, with

$$\delta V(r) = -Z\frac{\alpha^2}{3\pi}\int_{4m_e^2}^{\infty}\frac{dt}{t}\left(\frac{e^{-t^{1/2}r}}{r}\right)\delta\rho[t]. \quad (20)$$

This potential contributes to muonic-atom energies through the diagram shown in Fig. 4(a). Since Eq. (20) is a small perturbation and since the muon orbits of interest are appreciably larger in radius than the muon Compton wavelength, in evaluating matrix elements of $\delta V(r)$ we make the approximation of using nonrelativistic hydrogenic muon wave functions. [The same approximation applied to Eq. (19) yields the Uehling energy shifts for all of the levels in Table II to an accuracy of about 5%.[14]] Thus we take

$$R_{n\,n-1}(r) = \left[\left(\frac{2Z}{na_0}\right)^3 \frac{1}{(2n)!}\right]^{1/2} e^{-(Z/na_0)r}\left(\frac{2Zr}{na_0}\right)^{n-1},$$

$$a_0 = 1/\alpha m_\mu, \quad (21)$$

giving for the change in transition energy produced by $\delta V(r)$,

$$\delta E_\gamma = \delta E_n - \delta E_{n-1}$$
$$= \int_0^\infty r^2 dr[R_{n\,n-1}(r)^2 - R_{n-1\,n-2}(r)^2]\delta V(r). \quad (22)$$

Substituting Eq. (20) into Eq. (22), evaluating the r integral, and using $\alpha^2/(3\pi a_0)=4.35$ eV, we find

$$\delta\bar{E}_\gamma \equiv \frac{\delta E_\gamma}{4.35\ \mathrm{eV}\times Z^2[1/(n-1)^2-1/n^2]} = \int_{4m_e^2}^{\infty}\frac{dt}{t}f_\gamma[t]\delta\rho[t],$$

$$f_\gamma[t]=[1/(n-1)^2-1/n^2]^{-1}\times\left\{\frac{1}{(n-1)^2}\left[1+\left(\frac{t}{4m_\mu^2}\right)^{1/2}\frac{n-1}{Z\alpha}\right]^{-2(n-1)} - \frac{1}{n^2}\left[1+\left(\frac{t}{4m_\mu^2}\right)^{1/2}\frac{n}{Z\alpha}\right]^{-2n}\right\},$$

$$f_\gamma[0]=1. \tag{23}$$

Finally, for convenience in doing the numerical work we make the change of variable

$$t=4m_e^2e^w, \tag{24}$$

giving the formulas

$$\delta\bar{E}_\gamma=\int_0^\infty dw\, f_\gamma(w)\delta\rho(w),$$

$$f_\gamma(w)=f_\gamma[4m_e^2e^w], \tag{25}$$

$$\delta\rho(w)=\delta\rho[4m_e^2e^w].$$

Evidently, in the nonrelativistic approximation which we are using, the shifts in the transition energy δE_γ are j independent, and hence the two transitions for each n, l measure the same weighted integral of $\delta\rho(w)$. Thus, for purposes of comparison with Eq. (25) we average the two discrepancy values for each n, l, as shown in the fourth column of Table II.[15] The "reduced discrepancies" $\delta\bar{E}_\gamma$ introduced in Eq. (23) are tabulated in the final column of Table II.

Before proceeding further with our discussion of the muonic x-ray discrepancy, let us turn to consider another electrodynamic measurement which is sensitive to the asymptotic electronic vacuum polarization, the muon $g_\mu-2$ experiment. Here the conjectured deviation in the electronic vacuum-polarization spectral function contributes through the diagram of Fig. 4(b). Introducing the standard definition

$$a_\mu=\tfrac{1}{2}(g_\mu-2) \tag{26}$$

and using well-known formulas[16] for the photon spectral-function contribution to a, we find that changing the electron vacuum-polarization spectral function induces a $g_\mu-2$ discrepancy

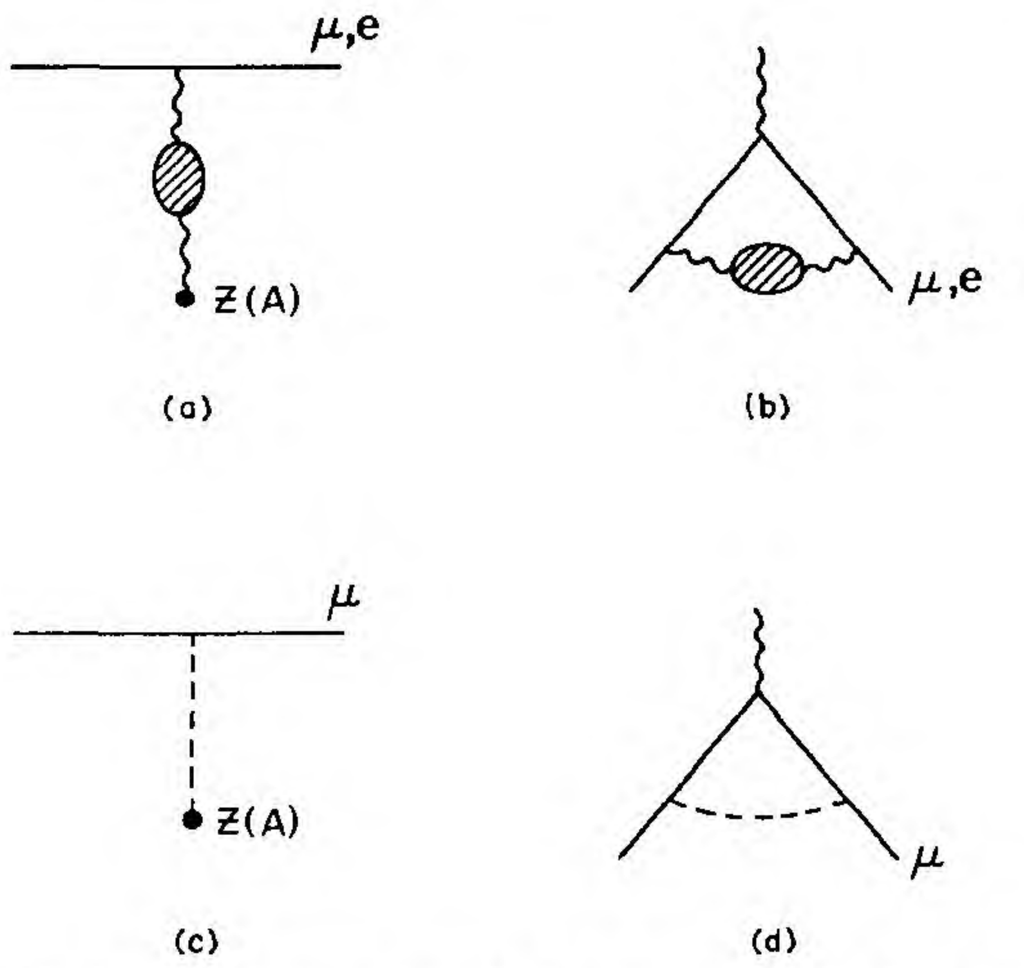


FIG. 4. (a) Diagram by which a vacuum-polarization modification (denoted by the shaded blob) contributes to μ- and e-atomic energy levels. (b) Diagram by which a vacuum-polarization modification contributes to $g_\mu-2$ and g_e-2. (c) Diagram by which a scalar-meson contributes to μ-atomic energy levels. (d) Diagram by which a scalar meson contributes to $g_\mu-2$.

$$\delta a_\mu=\frac{\alpha^2}{3\pi^2}\int_{4m_e^2}^{\infty}\frac{dt}{t}\frac{1}{2}f_a[t]\delta\rho[t],$$

$$f_a[t]=2\int_0^1 dx\frac{x^2(1-x)}{x^2+(1-x)t/m_\mu^2},\quad f_a[0]=1. \tag{27}$$

Using $\alpha^2/(3\pi^2)=1.80\times10^{-6}$ and making the change of variable of Eq. (24), we get the convenient formula

$$\delta a_\mu=1.80\times10^{-6}\tfrac{1}{2}\int_0^\infty dw\, f_a(w)\delta\rho(w),$$

$$f_a(w)=f_a[4m_e^2e^w]. \tag{28}$$

The result of carrying out the integrations in the expression for $f_a(t)$ is given in Appendix B.

Let us now return to our analysis of the muonic x-ray discrepancy. The kernels $f_\gamma(w)$ for four representative transitions are plotted in Fig. 5. Our numerical evaluation shows that the six transitions listed in Table III have weight functions f_γ which are nearly identical (their spread around curve b in Fig. 5 is less than one third of the spacing between curve b and curve a); averaging the weight functions for these transitions gives the function $\bar{f}_\gamma$ plotted in Fig. 6. Substituting the average of the reduced discrepancies for these six transitions into Eq. (25), we find

$$(54.5\pm10)\times10^{-3}=\text{average of six }(-\delta\bar{E}_\gamma) = -\int_0^\infty dw\,\bar{f}_\gamma(w)\delta\rho(w), \tag{29}$$

indicating that the sign of the discrepancy corresponds to a *reduction* in the electronic vacuum-polarization spectral function from its usual value of Eq. (19). Referring back to Fig. 6, we note that the function f_a is always greater than $\bar{f}_\gamma$. Hence if we assume that $\delta\rho(w)$ is always of negative sign in the region where f_a and $\bar{f}_\gamma$ are nonzero [as might reasonably be expected if we are just entering a new region of physics where the discrepancy $\delta\rho(w)$ is turning on], we learn that

$$\int_0^\infty dw\, f_a(w)\delta\rho(w) \leq -(54.5\pm 10)\times 10^{-3}. \qquad (30)$$

Comparing Eq. (30) with Eq. (28) we then get an inequality for the $g_\mu - 2$ discrepancy,

$$\delta\rho \leq 0 \Rightarrow \begin{cases} \delta a_\mu \leq -1.80\times 10^{-6}\tfrac{1}{2}(54.5\pm 10)\times 10^{-3} \\ \quad = -(0.49\pm 0.09)\times 10^{-7}, \\ \delta a_\mu/a_\mu \leq -42\pm 8 \text{ ppm}. \end{cases} \qquad (31)$$

A stronger prediction follows if, in addition to our assumption on the sign of $\delta\rho$, we assume that the magnitude of $\delta\rho$ increases monotonically with t (again as might reasonably be expected for an effect just turning on). Then defining

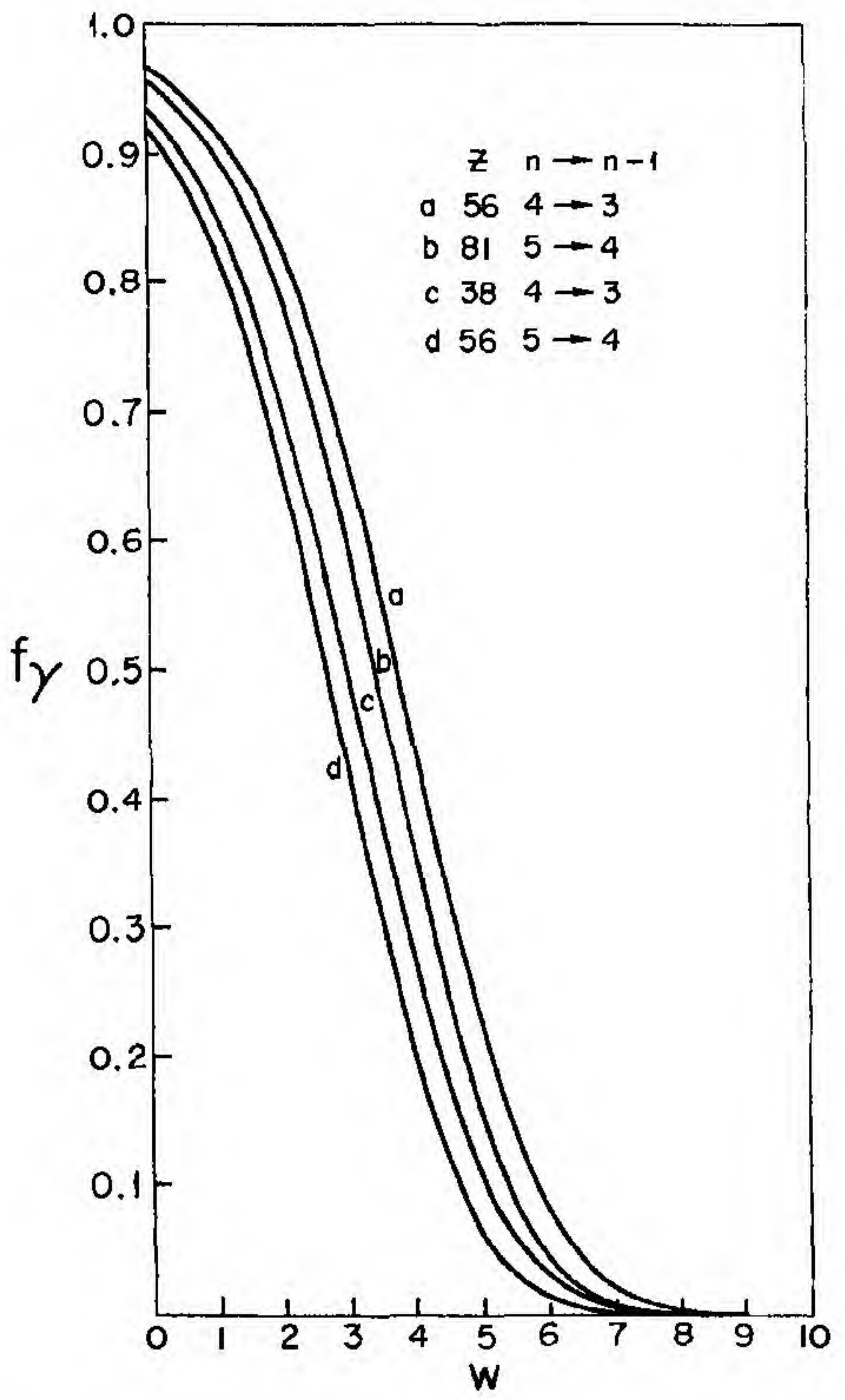


FIG. 5. Kernels f_γ for some representative transitions.

TABLE III. Transitions with nearly identical f_γ.

Element $_ZE$	Transition $^nl \to {}^{n-1}l-1$	Reduced average discrepancy $-\delta\bar{E}_\gamma$
$_{47}$Ag	$^4f \to {}^3d$	$(49.3\pm 30)\times 10^{-3}$
$_{48}$Cd	$^4f \to {}^3d$	$(20.5\pm 27)\times 10^{-3}$
$_{50}$Sn	$^4f \to {}^3d$	$(43.5\pm 26)\times 10^{-3}$
$_{80}$Hg	$^5g \to {}^4f$	$(71.8\pm 27)\times 10^{-3}$
$_{81}$Tl	$^5g \to {}^4f$	$(55.3\pm 26)\times 10^{-3}$
$_{82}$Pb	$^5g \to {}^4f$	$(73.7\pm 21)\times 10^{-3}$
Weighted average of six reduced discrepancies [a]:		$(54.5\pm 10)\times 10^{-3}$

[a] We have treated the errors as if they were purely statistical and have quoted the rms error for the average.

$$\bar{f}_\gamma(w) = 0, \quad w<0, \qquad (32)$$

we find that we can represent $f_a(w)$ as a superposition of displaced curves $\bar{f}_\gamma$,

$$f_a(w) = 1.016\bar{f}_\gamma(w) + \int_0^{10.2} dw'\, c(w')\bar{f}_\gamma(w-w'), \qquad (33)$$

with the positive weight function c plotted in Fig. 6. Multiplying by $\delta\rho(w)$ and integrating we get

$$\int_0^\infty dw\, f_a(w)\delta\rho(w) = 1.016\int_0^\infty dw\, \bar{f}_\gamma(w)\delta\rho(w) + \int_0^{10.2} dw'\, c(w') \times \int_0^\infty dw\, \bar{f}_\gamma(w-w')\delta\rho(w). \qquad (34)$$

But using Eq. (32) and the assumed monotonicity

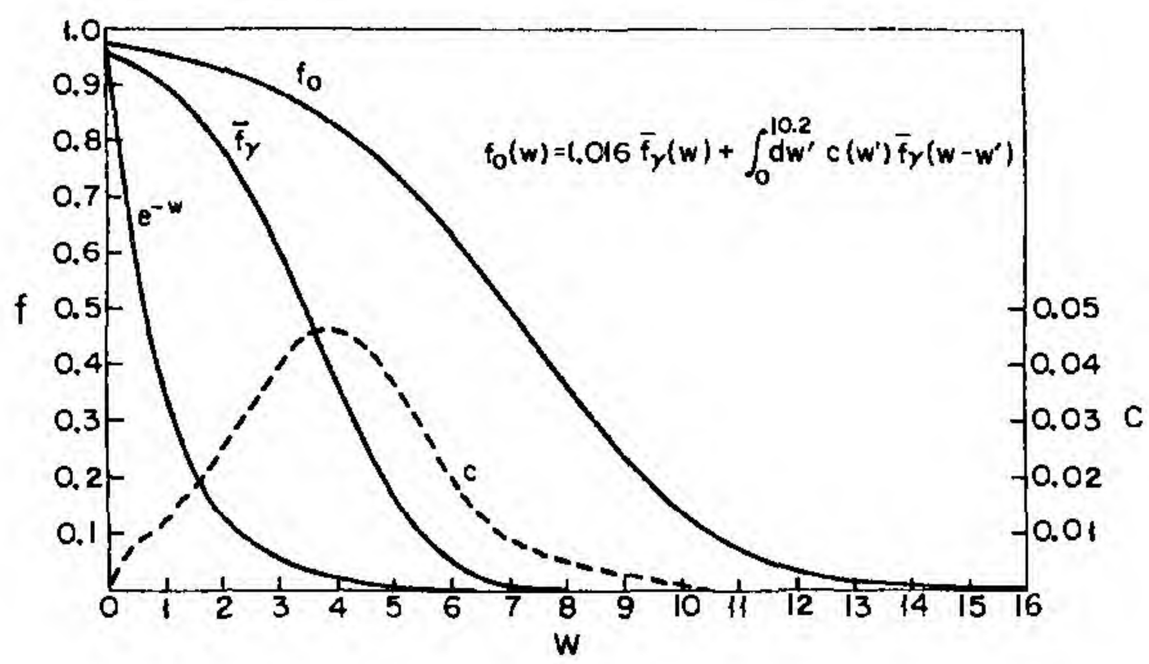


FIG. 6. Plots of the kernels $\bar{f}_\gamma$ and f_a [see the discussion which follows Eqs. (28) and (29) of the text].

of $\delta\rho$, we get

$$\int_0^\infty dw \bar{f}_\gamma(w-w')\delta\rho(w) = \int_0^\infty dw \bar{f}_\gamma(w)\delta\rho(w+w')$$

$$\leq \int_0^\infty dw \bar{f}_\gamma(w)\delta\rho(w) , \quad (35)$$

and so we learn

$$\int_0^\infty dw f_a(w)\delta\rho(w) \leq \left(1.016 + \int_0^{10.2} dw' c(w')\right) \times \int_0^\infty dw \bar{f}_\gamma(w)\delta\rho(w) = 2.00 \times \int_0^\infty dw \bar{f}_\gamma(w)\delta\rho(w) . \quad (36)$$

Thus adding the assumption of monotonicity doubles the prediction of Eq. (30), giving

$$\left.\begin{matrix}\delta\rho \leq 0 \\ |\delta\rho| \uparrow\end{matrix}\right\} \Rightarrow \left\{\begin{matrix}\delta a_\mu \leq -(0.98 \pm 0.18)\times 10^{-7} \\ \delta a_\mu/a_\mu \leq -84 \pm 16 \text{ ppm} .\end{matrix}\right. \quad (37)$$

Equation (37) is the principal result of our analysis.

Two remarks about Eq. (37) are in order. First, the discrepancy in a_μ predicted in Eq. (37) is compatible, within errors, with the present agreement of experiment with the conventional electrodynamic prediction for a_μ,[17]

$$a_\mu(\text{expt}) - a_\mu(\text{conventional QED}) = (2.5 \pm 3.1)\times 10^{-7} \quad (38)$$

However, it should be readily observable in the next $g_\mu - 2$ experiment, where it is anticipated[17] that the current experimental error of $\pm 3.1\times 10^{-7}$ ($=\pm 270$ ppm in $\delta a_\mu/a_\mu$) will be reduced by a factor of 20. Second, the predicted effect is substantially larger than the likely remaining uncertain contributions to a_μ. Specifically, these are the following.

(i) The 8th-order electrodynamic contribution to a_μ, which has been variously estimated[16] as 6×10^{-9}–7×10^{-9}, with an uncertainty of perhaps a few parts in 10^{-9}.

(ii) The uncertainty in the hadronic contribution to a_μ. Including the ρ, ω, and ϕ resonances and integrating the e^+e^- annihilation continuum up to $t_C = 25$ gives a known hadronic contribution of 71×10^{-9} with an estimated uncertainty of $\pm 7\times 10^{-9}$ (see Appendix B). The unknown contribution of the e^+e^- annihilation continuum beyond $t_C = 25$ will of course depend on the behavior of $R(t)$ in that region. To get a crude estimate, let us make the (hopefully extreme) assumption that $R(t)$ rises linearly up to $t = (460)^2$, where the one-photon annihilation cross section violates the $J = 1$ unitarity limit,[19] and cut off the integral at this point. This procedure suggests a bound on the high-energy hadronic contribution to a_μ of 15×10^{-9}. (Again see Appendix B.)

(iii) Unified gauge theories of the weak and electromagnetic interactions which do not have charged heavy leptons typically give contributions[20] to a_μ in the range from a few to ten parts in 10^{-9}. Specifically, the Weinberg-Salam SU(2)⊗ U(1) model predicts a contribution to a_μ of less than 9×10^{-9}. Thus, from (i), (ii), and (iii) we conclude that the sum of unknown contributions to a_μ is likely to be no bigger than $\approx 35\times 10^{-9}$, and hence should not mask the effect predicted in Eq. (37).

Although we have shown that the inequality of Eq. (37) does not contradict the current $g_\mu - 2$ experiment, we must still verify that it is possible to find specific functional forms $\delta\rho(w)$ which fit the muonic x-ray discrepancy without seriously violating any of the conventional tests of QED, including the very high precision $g_e - 2$ and Lamb-shift experiments.[21] A postulated vacuum-polarization discrepancy contributes to $g_e - 2$ through the diagram of Fig. 4(b), giving a formula identical to Eq. (27) apart from the replacement of m_μ in $f_a[t]$ by m_e. The smallness of m_e then permits use of the large-t asymptotic expression $\frac{1}{2}f_a \approx m_e^2/(3t)$, giving the simple expression

$$\delta a_e \approx 0.60\times 10^{-6} \int_{4m_e^2}^\infty \frac{dt}{t}\frac{m_e^2}{t}\delta\rho[t]$$

$$= 0.15\times 10^{-6} \int_0^\infty dw\, e^{-w}\delta\rho(w) . \quad (39)$$

Comparing Eq. (39) with the current difference between experiment and theory for a_e,[22]

$$a_e(\text{experiment}) - a_e(\text{conventional QED}) = (5.6 \pm 4.4)\times 10^{-9} , \quad (40)$$

we get the restriction

$$\int_0^\infty dw\, e^{-w}\delta\rho(w) = (37 \pm 29)\times 10^{-3} . \quad (41)$$

Next we consider the Lamb shift, which receives contributions from a vacuum-polarization discrepancy via the diagram of Fig. 4(a). Working again in the nonrelativistic hydrogenic approximation, we find for the change in the $2s$-$2p$ Lamb-transition energy

$$\delta\mathcal{L} \equiv \delta E_{2s\to 2p}$$
$$= \int_0^\infty r^2 dr [R_{20}(r)^2 - R_{21}(r)^2]\delta V(r)$$
$$= -\frac{a_0\alpha^2}{6\pi}\int_{4m_e^2}^\infty \frac{dt\,\delta\rho[t]}{(1+t^{1/2}a_0/Z)^4}. \quad (42)$$

Since $t^{1/2}a_0Z^{-1} = (t^{1/2}/m_e)\alpha^{-1}Z^{-1} \gg 1$, we can neglect the 1 in the denominator of Eq. (42), giving

$$\delta\mathcal{L} = -\frac{Z^4\alpha^5 m_e}{6\pi}\int_{4m_e^2}^\infty \frac{dt}{t}\frac{m_e^2}{t}\delta\rho[t], \quad (43)$$

which evidently measures the same integral over $\delta\rho$ as does $g_e - 2$. It is easy to see that the formula for the $ns - np$ Lamb transition is obtained by multiplying Eq. (43) by $(2/n)^3$. Hence, using the fact that

$$\frac{\alpha^5 m_e}{30\pi} \approx 27.1 \text{ MHz}, \quad (44)$$

we get the relation

$$\int_0^\infty dw\, e^{-w}\delta\rho(w)$$
$$= \frac{n^3[\mathcal{L}_{nZ}(\text{conventional QED}) - \mathcal{L}_{nZ}(\text{expt})]}{Z^4 \times 271 \text{ MHz}}. \quad (45)$$

In Table IV we have tabulated the right-hand side of Eq. (45) for a series of measured Lamb transitions.[23] Taking a weighted average of the four best determinations [the two measurements for H(n = 2) and the measurements for D(n = 2) and He$^+$(n = 2)], we find

$$\int_0^\infty dw\, e^{-w}\delta\rho(w) = (0.29 \pm 1.0)\times 10^{-3}, \quad (46)$$

evidently a much tighter restriction than is obtained from $g_e - 2$.

Our procedure for searching for satisfactory functional forms $\delta\rho$ is now as follows. Let

$$\begin{aligned}
&\delta\overline{E}_\gamma(i), \sigma(i) = \text{experimental reduced discrepancies and standard deviations from Table II, } i = 1,\ldots,12;\\
&F(i) = \text{theoretical fit to reduced discrepancies, } i = 1,\ldots,12;\\
&\delta a_\mu^{\text{th}} = \text{predicted change in } a_\mu,\\
&\delta I^{\text{th}} = \text{predicted value of } \int_0^\infty dw\, e^{-w}\delta\rho(w).
\end{aligned} \quad (47)$$

We form two χ^2:

$$\begin{aligned}
\chi^2_1 &\equiv \sum_{i=1}^{12}\left[\frac{F(i) - \delta\overline{E}_\gamma(i)}{\sigma(i)}\right]^2,\\
\chi^2_2 &= \chi^2_1 + \left(\frac{\delta a_\mu^{\text{th}} - 2.6\times 10^{-7}}{3.1\times 10^{-7}}\right)^2\\
&+ \left(\frac{\delta I^{\text{th}} - 37\times 10^{-3}}{29\times 10^{-3}}\right)^2\\
&+ \left(\frac{\delta I^{\text{th}} - 0.29\times 10^{-3}}{1.0\times 10^{-3}}\right)^2;
\end{aligned} \quad (48)$$

the first tests the fit to the muonic x-ray discrepancies alone, while the second tests the combined fit to the x-ray data and the $g_\mu - 2$, $g_e - 2$ and Lamb-shift experiments. For each assumed functional form of $\delta\rho$, we treat the over-all normalization as a free parameter and adjust it to minimize either χ^2_1 or χ^2_2, corresponding respectively to 12 − 1 = 11 or 15 − 1 = 14 degrees of freedom. A sampling of results of this procedure is shown in Tables V and VI. We conclude from these fits the following.[24]

(i) Functional forms giving good χ^2_2 fits can be found. When these same functional forms are fit by the χ^2_1 procedure the coefficients change by only about 25%, which is satisfactory.

(ii) The forms which give good χ^2_2 fits are all nearly step-function-like in character, with a turn-on at $w \approx 2$–3 [i.e., at $t \approx (30$–$80)m_e^2$]. The smallness below $w \sim 2$ is required by the Lamb-

TABLE IV. Lamb-shift measurements.

System	Conventional QED (MHz)	Expt (MHz)	Right-hand side of Eq. (41)
H(n = 2)	1057.911 ± 0.012	1057.90 ± 0.06	$(0.33 \pm 1.80)\times 10^{-3}$
		1057.86 ± 0.06	$(1.50 \pm 1.80)\times 10^{-3}$
H(n = 3)	314.896 ± 0.003	314.810 ± 0.052	$(8.57 \pm 5.18)\times 10^{-3}$
H(n = 4)	133.084 ± 0.001	133.18 ± 0.59	$(-22.7 \pm 139)\times 10^{-3}$
D(n = 2)	1059.271 ± 0.025	1059.28 ± 0.06	$(-0.27 \pm 1.92)\times 10^{-3}$
He$^+$ (n = 2)	14 044.765 ± 0.613	14 045.4 ± 1.2	$(-1.18 \pm 2.49)\times 10^{-3}$
He$^+$ (n = 3)	4184.42 ± 0.18	4183.17 ± 0.54	$(7.79 \pm 3.55)\times 10^{-3}$
He$^+$ (n = 4)	1769.088 ± 0.076	1769.4 ± 1.2	$(-4.60 \pm 17.7)\times 10^{-3}$
Li^{++} (n = 2)	62 762 ± 9	62 765 ± 21	$(-1.1 \pm 8.3)\times 10^{-3}$
C^{5+} (n = 2)	$(783.678 \pm 0.251)\times 10^{-3}$	$(744.0 \pm 7)\times 10^{-3}$	$(904 \pm 159)\times 10^{-3}$

TABLE V. Sample functional forms giving statistically satisfactory fits.

(1) Fits minimizing χ^2_2

Functional form $\delta\rho(w)$	χ^2_2	$10^7\delta a_\mu$	$\delta\mathcal{L}(H)$ (MHz) [a]
$-0.053\theta(w-3)$	12.1	-1.9	0.08
$-0.071\left(\frac{w-3}{w}\right)^{0.2}\theta(w-3)$	12.1	-2.1	0.08
$-0.16\left(\frac{w-2}{w}\right)^{2}\theta(w-2)$	12.9	-2.4	0.08

(2) Fits minimizing χ^2_1

Functional form $\delta\rho(w)$	χ^2_1	$10^7\delta a_\mu$	$\delta\mathcal{L}(H)$ (MHz) [a]
$-0.066\theta(w-3)$	7.9	-2.3	0.10
$-0.088\left(\frac{w-2}{w}\right)^{0.2}\theta(w-3)$	7.9	-2.6	0.10
$-0.21\left(\frac{w-2}{w}\right)^{2}\theta(w-2)$	8.1	-3.1	0.11

(3) Reduced discrepancies predicted by the fit $-0.071[(w-3)/w]^{0.2}\theta(w-3)$

Z	20	22	26	38	47	48	50	56	56	80	81	82
Transition $n \to n-1$	$3\to2$	$3\to2$	$3\to2$	$4\to3$	$4\to3$	$4\to3$	$4\to3$	$4\to3$	$5\to4$	$5\to4$	$5\to4$	$5\to4$
$-10^3\times\delta\bar{E}_\gamma$ (expt)	37.2 ±54	12.0 ±45	37.9 ±32	18.0 ±43	49.3 ±30	20.5 ±27	43.5 ±26	98.8 ±23	24.4 ±39	71.8 ±27	55.3 ±26	73.7 ±21
$-10^3\times\delta\bar{E}_\gamma$ (fit)	40.6	46.2	57.3	30.5	42.5	43.8	46.5	54.2	21.6	40.0	40.8	41.6

[a] See the comment in Ref. 25.

shift data, while the slow growth above the turn-on is needed in order not to violate the current limits on deviations in $g_\mu - 2$.

(iii) All of the good fits satisfy $\delta\rho \lesssim -0.03$ for large w. This is a general feature for any monotonic form $\delta\rho$ which is small in the Lamb-shift region $w \leq 2$, since (using the fact that $\bar{f}_\gamma \approx 0$ for $w \gtrsim 9$) we have

$$\begin{aligned} -54.5\times10^{-3} &= \int_0^\infty dw\, \bar{f}_\gamma(w)\delta\rho(w) \\ &\approx \int_2^9 dw\, \bar{f}_\gamma(w)\delta\rho(w) \\ &\gtrsim \delta\rho(9)\int_2^9 dw\, \bar{f}_\gamma(w) = 1.6\times\delta\rho(9), \end{aligned} \tag{49}$$

that is,

$$-0.034 \gtrsim \delta\rho(9). \tag{50}$$

Possible implications of Eq. (50) for QED tests involving timelike photon vertices will be discussed elsewhere.[25]

One additional place where a vacuum-polarization discrepancy should produce interesting effects is in the Lamb shift in muonic helium.[26] Applying Eq. (42) to this system (and noting that the $2p$ level here lies above the $2s$ level), we find

$$\begin{aligned} \delta\mathcal{L}([^4\mathrm{He},\mu]^+) &\equiv \delta E_{2p\to2s} \\ &= \frac{a_0\alpha^2}{6\pi}\int_{4m_e^2}^\infty dt \frac{\delta\rho[t]}{(1+t^{1/2}a_0/Z)^4} \\ &= \frac{2}{3\pi}\alpha\frac{m_e^2}{m_\mu}\int_0^\infty dw\, f_{\mathrm{He}}(w)\delta\rho(w), \end{aligned}$$

$$f_{\mathrm{He}}(w) = \frac{e^w}{[1+(m_e/m_\mu\alpha)e^{w/2}]^4}, \quad f_{\mathrm{He}}(0) \approx 0.13. \tag{51}$$

TABLE VI. Results of step-function fits.

Functional form $\delta\rho$	χ^2_2	$10^7\delta a_\mu$	$\delta\mathcal{L}(H)$ (MHz) [a]
$-0.004\theta(w-0.5)$	38	-0.23	0.08
$-0.014\theta(w-1.5)$	21	-0.69	0.10
$-0.032\theta(w-2.5)$	14	-1.3	0.09
$-0.072\theta(w-3.5)$	12	-2.3	0.07
$-0.16\theta(w-4.5)$	13	-4.0	0.06
$-0.37\theta(w-5.5)$	22	-6.7	0.05
$-0.39\theta(w-6.5)$	43	-5.0	0.02

[a] See the comment in Ref. 25.

Numerical evaluation of Eq. (51) shows that $f_{He}(w)/0.13$ lies within 20% of $\bar{f}_\gamma(w)$ in the range $0 \leq w \leq 6$ where neither is vanishingly small. Hence independent of the detailed form of $\delta\rho$, we find the prediction

$$\delta\mathcal{L}([^4\text{He}, \mu]^+) \sim \frac{2}{3\pi}\alpha\frac{m_e{}^2}{m_\mu} \times (-54.5\times10^{-3})\times 0.13$$

$$\approx -0.027 \text{ eV}, \qquad (52)$$

which may be an observable effect.

At this point let us conclude our examination of vacuum-polarization effects and turn to an alternative explanation for the muonic x-ray discrepancy, the possible existence[12] of a weakly coupled scalar, isoscalar boson ϕ. Interest in this explanation has been stimulated by the fact that such particles (with undetermined mass) are called for in unified gauge theories of the weak and electromagnetic interactions. Letting $g_{\phi\mu\bar{\mu}}$ and $g_{\phi N\bar{N}}$ denote the ϕ-muon and the ϕ-nucleon scalar couplings, and M_ϕ the ϕ mass, the potential produced by ϕ exchange between a muon and a nucleus of nucleon number A [Fig. 4(c)] is the simple Yukawa form[27]

$$V_\phi(r) = -\frac{g_{\phi\mu\bar{\mu}}g_{\phi N\bar{N}}}{4\pi}A\frac{e^{-M_\phi r}}{r}. \qquad (53)$$

Since a repulsive potential is required to remove the x-ray discrepancy, fitting Eq. (53) to the x-ray data will necessarily give $g_{\phi\mu\bar{\mu}}g_{\phi N\bar{N}} < 0$. As shown in Appendix C, this sign for the product of couplings is not possible in the simplest forms of gauge models, in which there is only one physical scalar meson and in which the chiral SU(3)⊗SU(3) symmetry-breaking term in the strong-interaction Lagrangian transforms as pure $(3,\bar{3})\oplus(\bar{3},3)$. Nonetheless, let us proceed in a purely phenomenological fashion and make a quantitative fit of Eq. (53) to the x-ray data. Replacing $\delta V(r)$ in Eq. (22) by $V_\phi(r)$, we find

$$\delta\bar{E}'_\gamma \equiv \frac{\delta\bar{E}_\gamma}{(A/2Z)} = 2.82\times10^4 g_{\phi\mu\bar{\mu}}\, g_{\phi N\bar{N}} f_\gamma[M_\phi{}^2]\ , \qquad (54)$$

with $\delta\bar{E}'_\gamma$ the "reduced discrepancy" appropriate to a potential which couples to A rather than to Z. The experimentally measured values of $\delta\bar{E}'_\gamma$ are tabulated in Table VII. Since in all gauge models the ϕ-electron coupling is expected to be of order $(m_e/m_\mu)g_{\phi\mu\bar{\mu}}$, the ϕ will have a negligible effect on the electron $g_e - 2$ and Lamb-shift measurements. So in fitting Eq. (54) to the data we minimize $\chi^2{}_1$ defined in Eq. (48), giving the results shown in Table VIII, in good agreement with the results quoted by Sundaresan and Watson.[28]

TABLE VII. Reduced discrepancies for scalar-meson calculation.

Element $_ZE$	Transition $^nl \to {}^{n-1}l-1$	Reduced average discrepancy $-\delta\bar{E}'_\gamma$
$_{20}$Ca	$^3d \to {}^2p$	$(37.2\pm54)\times10^{-3}$
$_{22}$Ti	$^3d \to {}^2p$	$(11.0\pm41)\times10^{-3}$
$_{26}$Fe	$^3d \to {}^2p$	$(35.1\pm30)\times10^{-3}$
$_{38}$Sr	$^4f \to {}^3d$	$(15.5\pm37)\times10^{-3}$
$_{47}$Ag	$^4f \to {}^3d$	$(42.9\pm26)\times10^{-3}$
$_{48}$Cd	$^4f \to {}^3d$	$(17.5\pm23)\times10^{-3}$
$_{50}$Sn	$^4f \to {}^3d$	$(36.6\pm22)\times10^{-3}$
$_{56}$Ba	$^4f \to {}^3d$	$(81.0\pm19)\times10^{-3}$
	$^5g \to {}^4f$	$(20.0\pm32)\times10^{-3}$
$_{80}$Hg	$^5g \to {}^4f$	$(57.0\pm21)\times10^{-3}$
$_{81}$Tl	$^5g \to {}^4f$	$(43.9\pm21)\times10^{-3}$
$_{82}$Pb	$^5g \to {}^4f$	$(58.5\pm17)\times10^{-3}$

Since a light scalar boson, as well as a vacuum-polarization anomaly, can satisfactorily fit the x-ray discrepancy, let us examine ways of distinguishing between the two possible explanations. First we consider the muonic-helium Lamb shift. Since $f_{He}(w) \approx \bar{f}_\gamma(w)$ for $0 \leq w \leq 6$, a scalar boson in the mass range from 1 to 22 MeV predicts an effect within about 20% of -0.027 eV, while for scalar bosons lighter than 1 MeV (corresponding to $w < 0$), the muonic-helium Lamb shift decreases as

$$\delta\mathcal{L}([^4\text{He}, \mu]^+) \sim \frac{-0.178 M_\phi{}^2}{(1+0.65M_\phi)^4}\text{ eV}, \qquad (55)$$

where M_ϕ is in MeV. Hence the muonic-helium experiment could only distinguish between a very light scalar boson[29] and the joint possibilities of a heavier scalar boson or a vacuum-polarization effect. On the other hand, the muonic vertex correction involving scalar-meson exchange makes a small positive[20] contribution to a_μ, as distinct from the sizable negative contribution predicted by a vacuum-polarization anomaly. So the next generation of $g_\mu - 2$ experiments should unambiguously distinguish between the vacuum-

TABLE VIII. Results of scalar-meson fits.

M_ϕ (MeV)	$\chi^2{}_1$	$(g_{\phi\mu\bar{\mu}}g_{\phi N\bar{N}})/4\pi$
0.5	8.1	-1.3×10^{-7}
1	7.9	-1.4×10^{-7}
4	6.8	-2.0×10^{-7}
8	6.1	-3.8×10^{-7}
12	6.5	-6.9×10^{-7}
16	7.5	-1.2×10^{-6}
22	10.1	-2.5×10^{-6}

polarization and scalar-meson explanations for the muonic-atom x-ray discrepancy.

ACKNOWLEDGMENTS

I wish to thank M. Baker, K. Johnson, and, above all, S. Brodsky for conversations which stimulated the muonic x-ray part of this paper. I have benefited from conversations with J. Bahcall, L. S. Brown, R. F. Dashen, B. Lautrup, B. W. Lee, J. Rafelski, V. Telegdi, S. B. Treiman, W. J. Weisberger, T.-M. Yan, and A. Zee, and wish to thank S. B. Treiman for reading the manuscript.

APPENDIX A: VECTOR-MESON PARAMETERS

For the ω and ϕ vector-meson parameters we take[30]

$$M_\omega = 784 \text{ MeV}, \quad \Gamma(\omega\to e^+e^-) = 0.76 \text{ keV},$$
$$M_\phi = 1019 \text{ MeV}, \quad \Gamma(\phi\to e^+e^-) = 1.36 \text{ keV}. \tag{A1}$$

For $F_\pi(t)$ we use the Gounaris-Sakurai formula[31] with an $\omega\to 2\pi$ interference term,[30]

$$F_\pi(t) = \frac{M_\rho^2(1+\delta\Gamma_\rho/M_\rho)}{M_\rho^2 - t + H(t) - iM_\rho\Gamma_\rho(k/k_\rho)^3 M_\rho/\sqrt{t}} + Ae^{i\alpha}\frac{M_\omega^2}{M_\omega^2 - t - iM_\omega\Gamma_\omega},$$
$$H(t) = \frac{\Gamma_\rho M_\rho^2}{k_\rho^3}\{k^2[h(t) - h(M_\rho^2)] + k_\rho^2 h'(M_\rho^2)(M_\rho^2 - t)\},$$
$$h(t) = \frac{2}{\pi}\frac{k}{\sqrt{t}}\ln\left(\frac{\sqrt{t}+2k}{2m_\pi}\right), \tag{A2}$$
$$h'(M_\rho^2) = \frac{1}{2\pi M_\rho^2} + \frac{m_\pi^2}{\pi M_\rho^3 k_\rho}\ln\left(\frac{M_\rho+2k_\rho}{2m_\pi}\right),$$
$$k = (\tfrac{1}{4}t - m_\pi^2)^{1/2}, \quad k_\rho = (\tfrac{1}{4}M_\rho^2 - m_\pi^2)^{1/2},$$
$$A = \frac{6B^{1/2}(\omega - ee)\Gamma_\omega}{\alpha M_\omega \beta_\pi^{3/2}} B^{1/2}(\omega\to 2\pi),$$
$$\beta_\pi = \left(1 - \frac{4m_\pi^2}{t}\right)^{1/2},$$

with the following values for the parameters[30]:

$$\delta = 0.6, \quad \alpha = 86^\circ,$$
$$M_\rho = 775 \text{ MeV}, \quad \Gamma_\omega = 9.2 \text{ MeV},$$
$$\Gamma_\rho = 149 \text{ MeV}, \quad B^{1/2}(\omega\to ee) = 0.906\times 10^{-2}, \tag{A3}$$
$$m_\pi = 140 \text{ MeV}, \quad B^{1/2}(\omega\to 2\pi) = 0.19.$$

As discussed in Appendix B, approximating the small-t region in this fashion as a sum of ω, ϕ, and ρ contributions should yield the small-t contribution to $T_{\rm obs}(-s)$ (which is only a small fraction of the total for large $-s$) to an accuracy of about 15%.

APPENDIX B: FORMULAS FOR $f_a[t]$ AND THE HADRONIC CONTRIBUTION TO $g_\mu - 2$

The function $f_a[t]$ appearing in Eq. (27) has been evaluated by Brodsky and de Rafael,[32] who find

$$f_a[t] = 2K(t),$$
$$0\le t\le 4m_\mu^2, \quad \tau = t/4m_\mu^2,$$
$$K(t) = \tfrac{1}{2} - 4\tau - 4\tau(1-2\tau)\ln(4\tau) - 2(1-8\tau+8\tau^2)\left(\frac{\tau}{1-\tau}\right)^{1/2}\arctan\left(\frac{1-\tau}{\tau}\right)^{1/2};$$
$$t\ge 4m_\mu^2, \quad x = \frac{1-(1-4m_\mu^2/t)^{1/2}}{1+(1-4m_\mu^2/t)^{1/2}},$$
$$K(t) = \tfrac{1}{2}x^2(2-x^2) + (1+x)^2(1+x^2)\frac{\ln(1+x) - x + \frac{1}{2}x^2}{x^2} + \frac{1+x}{1-x}x^2\ln x. \tag{B1}$$

Corresponding to the division of $T_{\rm obs}(-s)$ into four pieces in Eq. (15), we write the hadronic contribution to a_μ as

$$a_\mu = \frac{1}{4\pi^3}\int_{4m_\pi^2}^\infty dt\,\sigma(e^+e^-\to\text{hadrons}; t)K(t)$$
$$= a_\mu^{\omega+\phi} + a_\mu^\rho + a_\mu^{\text{cont}(1)} + a_\mu^{\text{cont}(2)}. \tag{B2}$$

Working in the same narrow-resonance approximation as in the text, we find for $a_\mu^{\omega+\phi}$ the expression[5]

$$a_\mu^{\omega+\phi} = \sum_{V=\omega,\phi}\frac{3}{\pi}K(M_V^2)\frac{\Gamma(V\to e^+e^-)}{M_V}, \tag{B3}$$

while a_μ^ρ is given by the integral

$$a_\mu^\rho = \frac{\alpha^2}{12\pi^2}\int_{4m_\pi^2}^\infty\frac{dt}{t}\left(1 - \frac{4m_\pi^2}{t}\right)^{3/2}|F_\pi(t)|^2 K(t). \tag{B4}$$

Substituting the parameters from Appendix A and evaluating Eqs. (B3) and (B4) numerically gives

$$a_\mu^{\omega+\phi} = 9.1\times 10^{-9}, \quad a_\mu^\rho = 45\times 10^{-9},$$
$$a_\mu^{\text{tot small }t} = 54\times 10^{-9}. \tag{B5}$$

A more elaborate evaluation of the small-t contribution has been given by Bramôn, Etim, and Greco,[33] who sum the contributions of the various important hadronic states directly from Eq. (B2), giving

$$a_\mu^{\text{tot small }t} = (61\pm 7)\times 10^{-9}, \tag{B6}$$

indicating that our method of treating the small-t region is good to about 15%. To evaluate $a_\mu^{\text{cont}(1)}$

and $a_\mu^{\rm cont\,(2)}$ we approximate $K(t)$ by its asymptotic form

$$K(t) \approx \tfrac{1}{3} m_\mu{}^2/t, \quad t \to \infty, \tag{B7}$$

giving [in units where unity = $(1\ \mathrm{GeV}/c)^2$]

$$\begin{aligned} a_\mu^{\rm cont\,(1)} &= 6.7\times 10^{-9} \int_{0.39}^{25} \frac{dt R(t)}{t^2}, \\ a_\mu^{\rm cont\,(2)} &= 6.7\times 10^{-9} \int_{25}^{t_C} \frac{dt R(t)}{t^2}. \end{aligned} \tag{B8}$$

Evaluating $a_\mu^{\rm cont\,(1)}$ numerically using the data plotted in Fig. 1 gives

$$a_\mu^{\rm cont\,(1)} = 9.6 \pm 2, \tag{B9}$$

with the error a rough guess. Thus the total known hadronic contribution to a_μ is

$$(61\pm 7)\times 10^{-9} + (9.6\pm 2)\times 10^{-9} = (71 \pm 7)\times 10^{-9}. \tag{B10}$$

Estimating the unmeasured contribution by assuming a linearly rising $R(t)$ up to $t_C = (2.230)^2$, we get

$$\begin{aligned} a_\mu^{\rm cont\,(2)} &= 6.7\times 10^{-9} \times 0.24 \times \ln\left(\frac{(460)^2}{25}\right) \\ &= 15 \times 10^{-9}, \end{aligned} \tag{B11}$$

as stated in the text.

APPENDIX C: SIGN OF THE SCALAR-MESON EXCHANGE POTENTIAL IN SIMPLE GAUGE THEORIES

Consider a gauge theory of the weak and electromagnetic interactions in which only one scalar field ϕ develops a vacuum expectation, $\phi \to \phi + \lambda$, as a result of spontaneous symmetry breaking. Since λ is the source of the lepton masses, the interaction Hamiltonian (= − the interaction Lagrangian) coupling ϕ to the muons is[34]

$$\mathcal{H}_{\phi\mu\bar{\mu}} = \frac{\phi}{\lambda} m_\mu \bar{\psi}_\mu \psi_\mu . \tag{C1}$$

Since in the hadronic sector λ is the origin of chiral SU(3)⊗ SU(3) symmetry breaking, the interaction Hamiltonian coupling ϕ to the hadrons is[34]

$$\mathcal{H}_{\phi\,\mathrm{hadron}} = \frac{\phi}{\lambda} \delta\mathcal{H}_{\mathrm{chiral\ breaking}} . \tag{C2}$$

Hence the sign of $g_{\phi\mu\bar{\mu}} g_{\phi N\bar{N}}$ is the same as the sign of $\langle N | \delta\mathcal{H}_{\mathrm{chiral\ breaking}} | N\rangle$. Now if $\delta\mathcal{H}_{\mathrm{chiral\ breaking}}$ transforms under SU(3)⊗ SU(3) as $(3,\bar{3})\oplus(\bar{3},3)$, then using the notation of Gell-Mann, Oakes, and Renner[35] we readily find that

$$\begin{aligned} \langle N | \delta\mathcal{H}_{\mathrm{chiral\ breaking}} | N\rangle &= \frac{3\sigma_{\pi NN}}{(\sqrt{2}+c)\sqrt{2}} + \left(c - \frac{1}{\sqrt{2}}\right)\langle N | u_8 | N\rangle, \\ c &= -\sqrt{2}\frac{m_K{}^2 - m_\pi{}^2}{m_K{}^2 + \frac{1}{2} m_\pi{}^2} = -1.25, \\ \langle N | u_8 | N\rangle &= \text{baryon mass splitting parameter} \\ &= 170\ \mathrm{MeV}, \\ \sigma_{\pi NN} &\equiv \tfrac{1}{3}(\sqrt{2}+c)\langle N | \sqrt{2} u_0 + u_8 | N\rangle . \end{aligned} \tag{C3}$$

That is, we have

$$\langle N | \delta\mathcal{H}_{\mathrm{chiral\ breaking}} | N\rangle = 12.9\sigma_{\pi NN} - 333\ \mathrm{MeV}. \tag{C4}$$

Recent determinations of the σ term $\sigma_{\pi NN}$ suggest a value in the range 45–85 MeV,[36] making $\langle N | \delta\mathcal{H}_{\mathrm{chiral\ breaking}} | N\rangle$ positive and giving an attractive scalar-meson exchange force. A value of $\sigma_{\pi NN}$ smaller than 25 MeV would be needed to make the scalar-meson exchange force repulsive, as is required to explain the muonic x-ray discrepancy.

---

[1] For a summary of the recent data, see R. Gatto and G. Preparata, Phys. Lett. 50B, 479 (1974).

[2] A good review of the relevant experiments and theory is given by P. J. S. Watson and M. K. Sundaresan, Can. J. Phys. 52, 2037 (1974).

[3] L. Silvestrini, in *Proceedings of the XVI International Conference on High Energy Physics, Chicago-Batavia, Ill., 1972*, edited by J. D. Jackson and A. Roberts (NAL, Batavia, Ill., 1973), Vol. 4, p. 1.

[4] T. Appelquist and H. Georgi, Phys. Rev. D 8, 4000 (1973); A. Zee, *ibid.* 8, 4038 (1973).

[5] See, for example, C. H. Llewellyn Smith, CERN Report No. CERN TH. 1849, 1974 (unpublished); N. Cabibbo and G. Karl, CERN Report No. CERN TH. 1858, 1974 (unpublished).

[6] We follow the treatment of the vector-meson contributions to $\sigma(e^+e^- \to \mathrm{hadrons};t)$ given by M. Gourdin and E. de Rafael, Nucl. Phys. B10, 667 (1969).

[7] J. Arafune, Phys. Rev. Lett. 32, 560 (1974); L. S. Brown, R. N. Cahn, and L. D. McLerran, *ibid.* 32, 562 (1974); M. Gyulassy, *ibid.* 32, 1393 (1974).

[8] J. Rafelski, B. Müller, G. Soff, and W. Greiner, Univ. of Pennsylvania report (unpublished).

[9] M. S. Dixit *et al.*, Phys. Rev. Lett. 27, 878 (1971).

[10] H. Walter *et al.*, Phys. Lett. 40B, 197 (1972).

[11] S. L. Adler, Phys. Rev. D 5, 3021 (1972); K. Johnson and M. Baker, *ibid.* 8, 1110 (1973).

[12] M. K. Sundaresan and P. J. S. Watson, Phys. Rev. Lett. 29, 15 (1972); L. Resnick, M. K. Sundaresan, and

P. J. S. Watson, Phys. Rev. D 8, 172 (1973).

[13]To make the definition of $\delta\rho$ precise, we are assuming $\delta\rho$ to be an extra contribution to the vacuum-polarization two-point spectral function above and beyond the usual second- and fourth-order electron and muon contributions, of a magnitude much larger than the naive perturbative estimate of the sixth- and higher-order terms. Because $Z\alpha$ is not a very small parameter for some of the atomic species being considered, in establishing the existence of the muonic anomaly it is important to take into account 4, 6, . . . -point vacuum-polarization amplitudes in which one vertex emits a photon coupling to the muon and all the remaining vertices couple to the nuclear Coulomb potential. Numerically, the contribution of such diagrams to the x-ray energies is only a few percent of the Uehling potential contribution, so we neglect possible nonperturbative modifications of the higher-point functions.

[14]Just to set the energy scale involved, the Uehling energies are typically 0.2–3 keV out of total transition energies of 150–500 keV.

[15]We assume independence of errors and add erros in quadrature for both the muonic discrepancies and the Lamb-shift measurements. If systematic errors are large, then the errors for average quantities will be larger than those quoted, making the constraints on the χ^2 fits less restrictive than those we have assumed.

[16]B. Lautrup, A. Petermann, and E. de Rafael, Phys. Rep. 3C, 193 (1972).

[17]V. W. Hughes, in *Atomic Physics 3*, edited by S. J. Smith and G. K. Walters (Plenum, New York, 1973), p. 1; B. Lautrup (unpublished).

[18]B. Lautrup, Phys. Lett. 38B, 408 (1972); M. A. Samuel, Phys. Rev. D 9, 2913 (1974).

[19]C. H. Llewellyn Smith, Ref. 5; N. Cabibbo and R. Gatto, Phys. Rev. 124, 1577 (1961).

[20]R. Jackiw and S. Weinberg, Phys. Rev. D 5, 2396 (1972); I. Bars and M. Yoshimura, *ibid.* 6, 374 (1972); J. R. Primack and H. R. Quinn, *ibid.* 6, 3171 (1972).

[21]The atomic fermium K_α x-ray measurements of P. F. Dittner *et al.* [Phys. Rev. Lett. 26, 103 (1971)] only probe the Uehling potential to an accuracy of about 20%. See B. Fricke *et al.*, Phys. Rev. Lett. 28, 714 (1972). Constraints imposed by timelike photon vertex measurements are discussed in S. L. Adler, R. F. Dashen, and S. B. Treiman, following paper, Phys. Rev. D 10, 3728 (1974).

[22]N. M. Kroll, in *Atomic Physics 3*, edited by S. J. Smith and G. K. Walters (Plenum, New York, 1973), p. 33.

[23]The data are taken from Ref. 16, except for Li^{++}, where we use the more accurate measurements of P. Leventhal and P. G. Havey, Phys. Rev. Lett. 32, 808 (1974).

[24]The conclusion that satisfactory fits can be found has also been reached by F. Heile (unpublished), using a momentum-space parametrization of the vacuum-polarization discrepancy.

[25]S. L. Adler, R. F. Dashen, and S. B. Treiman (in preparation). We show in this paper that a decrease in the vacuum-polarization spectral function necessarily implies a decrease in the vertex for emission of a timelike photon, and tests of the effect are suggested. Obviously, modifications in vertex parts will at some level make contributions to the electrodynamic processes discussed in the present paper above and beyond the direct effect of the postulated vacuum-polarization discrepancy. There seems at present to be no way of estimating the size of such additional contributions; about all one can say is that they are likely to be most important in the electron Lamb shift and g_e-2 experiments, where only one mass scale (m_e) is involved and the vacuum-polarization and photon-electron vertex parts are off-shell to a similar degree. Hence the electronic Lamb-shift predictions of Tables V and VI should not be taken too literally. On the other hand, in the muon energy level and $g_\mu-2$ experiments, two mass scales (both m_e and m_μ) are involved, with the electron vacuum-polarization loops much further off-shell (relative to their natural mass scale) than are the photon-muon vertices. Thus in this case the neglect of possible vertex modifications, which is implicit in all of the discussion of the text, may well be justified.

[26]E. Campini, Lett. Nuovo Cimento 4, 982 (1970); P. J. S. Watson and M. K. Sundaresan, Ref. 2. As both of these references emphasize, in order for muonic helium to be useful for electrodynamics tests, current uncertainties in the helium nuclear charge radius and nuclear polarizability will have to be reduced.

[27]The factor $(4\pi)^{-1}$ in Eq. (49) appears to have been omitted in the basic paper of R. Jackiw and S. Weinberg, Phys. Rev. D 5, 2396 (1972) and in subsequent papers quoting their formulas.

[28]Because of the factor of 4π mentioned in Ref. 26, our $(g_{\phi\mu\bar{\mu}}g_{\phi N\bar{N}})/4\pi$ should be compared with $g_{\phi\mu\bar{\mu}}g_{\phi N\bar{N}}$ of Ref. 2. Omitting the $_{20}$Ca, $_{22}$Ti, $_{26}$Fe, and $_{38}$Sr discrepancies from the fit, as was done in Ref. 2, we find an effective coupling of -7.6×10^{-7} at $M_\phi=12$ MeV, in agreement with the magnitude of 8.0×10^{-7} quoted in Ref. 2. The numerical results of Table VIII were obtained by fitting to all discrepancy data.

[29]The possibility of a very light scalar meson may well be almost academic. An experiment reported by D. Kohler, J. A. Becker, and B. A. Watson [Phys. Rev. Lett. 33, 1628 (1974)] looks, via the e^+e^- decay mode, for a ϕ produced in the transition from ^{16}O(6.05 MeV) and ^{4}He(20.2 MeV) 0^+ states to the 0^+ ground states, and concludes that M_ϕ cannot be between 1.030 MeV and 18.2 MeV. Furthermore, neutron-electron scattering data rule out $M_\phi \lesssim 0.6$ MeV (see Ref. 24), leaving only a narrow allowed region between 0.6 and 1.03 MeV. These remarks do not apply to the derivative-coupled ϕ discussed recently by S. Barshay (unpublished), where the electron coupling is smaller than the μ coupling by two, as opposed to one, powers of m_e/m_μ.

[30]The data used are taken from J. Lefrançois, in *Proceedings of the 1971 International Symposium on Electron and Photon Interactions at High Energies*, edited by N. B. Mistry (Laboratory of Nuclear Studies, Cornell University, Ithaca, N. Y., 1972), p. 51.

[31]G. J. Gounaris and J. J. Sakurai, Phys. Rev. Lett. 21, 244 (1968).

[32]S. J. Brodsky and E. de Rafael, Phys. Rev. 168, 1620 (1968); 174, 1835 (1968).

[33]A. Bramõn, E. Etim, and M. Greco, Phys. Lett. 39B, 514 (1972).

[34]R. Jackiw and S. Weinberg, Ref. 20.

[35]M. Gell-Mann, R. J. Oakes, and B. Renner, Phys. Rev.

175, 2195 (1968). See also B. Renner, in *Springer Tracts in Modern Physics*, edited by G. Höhler and E. A. Niekisch (Springer, New York, 1972), Vol. 61, p. 121.

[36]H. Pilkuhn *et al.*, Nucl. Phys. B65, 460 (1973). See especially pp. 480–481. The estimate of Eq. (C4) is also given by E. Reya, Rev. Mod. Phys. 46, 545 (1974). Reya writes $M_B = M_0 + \langle B|\delta\mathcal{H}_{\text{chiral breaking}}|B\rangle$, where M_0 is the baryon mass in the absence of SU(3) $\otimes$ SU(3) breaking. For the $(3,\bar{3})\oplus(\bar{3},3)$ case, he finds $M_0 = 1300\ \text{MeV} - 13\sigma_{\pi NN}$, so for $\sigma_{\pi NN}$ in the range 45–85 MeV the mass M_0 is less than the nucleon mass, making $\langle N|\delta\mathcal{H}_{\text{chiral breaking}}|N\rangle$ positive.

PHYSICAL REVIEW D VOLUME 9, NUMBER 1 1 JANUARY 1974

$I=\frac{1}{2}$ contributions to $\nu_\mu + N \to \nu_\mu + N + \pi^0$ in the Weinberg weak-interaction model

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540
and National Accelerator Laboratory, Batavia, Illinois 60510*

(Received 8 August 1973)

We use a detailed dispersion-theoretic model for pion production in the (3, 3)-resonance region to calculate the ratio

$$R=\frac{\sigma(\nu_\mu+n\to\nu_\mu+n+\pi^0)+\sigma(\nu_\mu+p\to\nu_\mu+p+\pi^0)}{2\sigma(\nu_\mu+n\to\mu^-+p+\pi^0)}$$

in the Weinberg weak-interaction theory. We find that $I=\frac{1}{2}$ contributions do not substantially modify the earlier static model calculation of R given by B. W. Lee.

Neutral-pion production by neutrinos appears to be one of the best reactions for searching for the hadronic weak neutral current predicted by the Weinberg weak-interaction theory.[1] In fact, if one accepts the bound

$$\sin^2\theta_W \leq 0.35 \tag{1}$$

given by Gurr, Reines, and Sobel,[2] the static-model calculation by B. W. Lee[3] of

$$R=\frac{\sigma(\nu_\mu+n\to\nu_\mu+n+\pi^0)+\sigma(\nu_\mu+p\to\nu_\mu+p+\pi^0)}{2\sigma(\nu_\mu+n\to\mu^-+p+\pi^0)} \tag{2}$$

in the Weinberg theory is already in conflict with existing experiments[4] in complex nuclei. Two essential cautions are necessary, however, before concluding that the Weinberg theory is ruled out. First, charge-exchange effects are important in complex nuclei, and may result in an experimentally measured value of R which is smaller than the true single-nucleon-target value by a factor of up to 2.[5] Second, the static-model approximation, which neglects $I=\frac{1}{2}$ s-channel contributions to the reactions in Eq. (2), has the effect of overestimating R.[6] If the $I=\frac{1}{2}$ corrections are large enough, then, together with charge-exchange corrections, they may move experiment and theory back into agreement.

In this note we report the results of calculating the $I=\frac{1}{2}$ corrections to R using the detailed dispersion-theoretic model of weak pion production in the (3, 3)-resonance region which we developed some time ago.[7] The model is basically a generalization to weak pion production of the old CGLN model for pion photoproduction.[8] Nonresonant multipoles are treated in the Born approximation,[9] while the resonant (3, 3)-channel multipoles are obtained from the Born approximation by a unitarization procedure. The model is in excellent agreement with pion photoproduction data,[7] agrees well with pion electroproduction data up to a four-momentum transfer of $k^2\approx 0.5$ (GeV/c)^2,[7] and is also in satisfactory accord with the recent Argonne measurements of weak pion production.[10] Because all terms contributing to the weak-production amplitude in the model are proportional to nucleon elastic form factors, the model fails badly in the region $k^2 \gg 0.5$ (GeV/c)^2, where scaling effects become visible and leptonic inelastic cross sections decrease more slowly with increasing k^2 than elastic form factors squared. Fortunately, this region of large k^2 makes a relatively small contribution to the individual cross sections in Eq. (2), and the errors will furthermore tend to cancel between numerator and denominator.

In the Weinberg model, the effective Lagrangian for the semileptonic strangeness-conserving weak interactions is

$$\mathcal{L}=\frac{G}{\sqrt{2}}\cos\theta_C\{\bar{\mu}\gamma_\lambda(1+\gamma_5)\nu_\mu(J_\lambda^{V1}+iJ_\lambda^{V2}+J_\lambda^{A1}+iJ_\lambda^{A2}) + \bar{\nu}_\mu\gamma_\lambda(1+\gamma_5)\nu_\mu[J_\lambda^{V3}(1-2\sin^2\theta_W)+J_\lambda^{A3}-2\sin^2\theta_W J_\lambda^S]+\cdots\}, \tag{3}$$

where we have shown both the charged- and the neutral-current terms contributing to Eq. (2). In terms of the isospin matrix elements defined in Eqs. (2B.4) and (2B.5) of Ref. 7, the hadronic matrix element of the neutral current is

$$_{\rm out}\langle \pi N|J_\lambda^{V3}(1-2\sin^2\theta_W)+J_\lambda^{A3}-2\sin^2\theta_W J_\lambda^S|N\rangle=(1-2\sin^2\theta_W)[a_B^{(3/2)}V_\lambda^{(3/2)}+a_B^{(1/2)}V_\lambda^{(1/2)}] -2\sin^2\theta_W a_B^{(0)}V_\lambda^{(0)}+a_B^{(3/2)}A_\lambda^{(3/2)}+a_B^{(1/2)}A_\lambda^{(1/2)}. \tag{4}$$

The amplitudes appearing in Eq. (4) are all ones which appear in either the pion-electroproduction or the weak-production calculations of Ref. 7, and so R can be evaluated by a simple adaptation of the computer routines used in the earlier work. The result of such a calculation is shown in Fig. 1, where we have assumed an incident lab neutrino energy $k_{10}^{L}=1$ GeV and a nucleon axial-vector elastic form factor

$$g_A(k^2)=\frac{1.24}{[1+k^2/(1\ \mathrm{GeV}/c)^2]^2}, \tag{5}$$

and have integrated over the (3, 3)-resonance region up to a maximum isobar mass of $W=1.47$ GeV. Curve a gives the result obtained from our model when both resonant and nonresonant multipoles are kept; curve b is the corresponding result obtained when only the resonant multipoles are kept, and hence when $I=\frac{1}{2}$ amplitudes are neglected. As expected, curve a lies below curve b, but the effect of the $I=\frac{1}{2}$ corrections is not dramatic. For comparison, we give in curve c the result obtained from Lee's static-model calculation.[11] If one assumes that θ_W is restricted as in Eq. (1), and includes a safety factor of 2 for charge-exchange effects, curve a is barely consistent with the present experimental upper bounds. Put conservatively, our calculations indicate that an experiment to measure R at the level of a few percent should be decisive.

Note added in proof. Recently, the possible observation of neutral current events has been reported in deep-inelastic inclusive neutrino reactions by the CERN Gargamelle group.[12] If confirmed, this experiment will establish the *existence* of neutral currents; however, more detailed questions, such as whether the phenomenological form of Eq. (3) is correct, will require the independent study of many different neutral-current induced reactions, among them the pion-production reaction considered in this note.

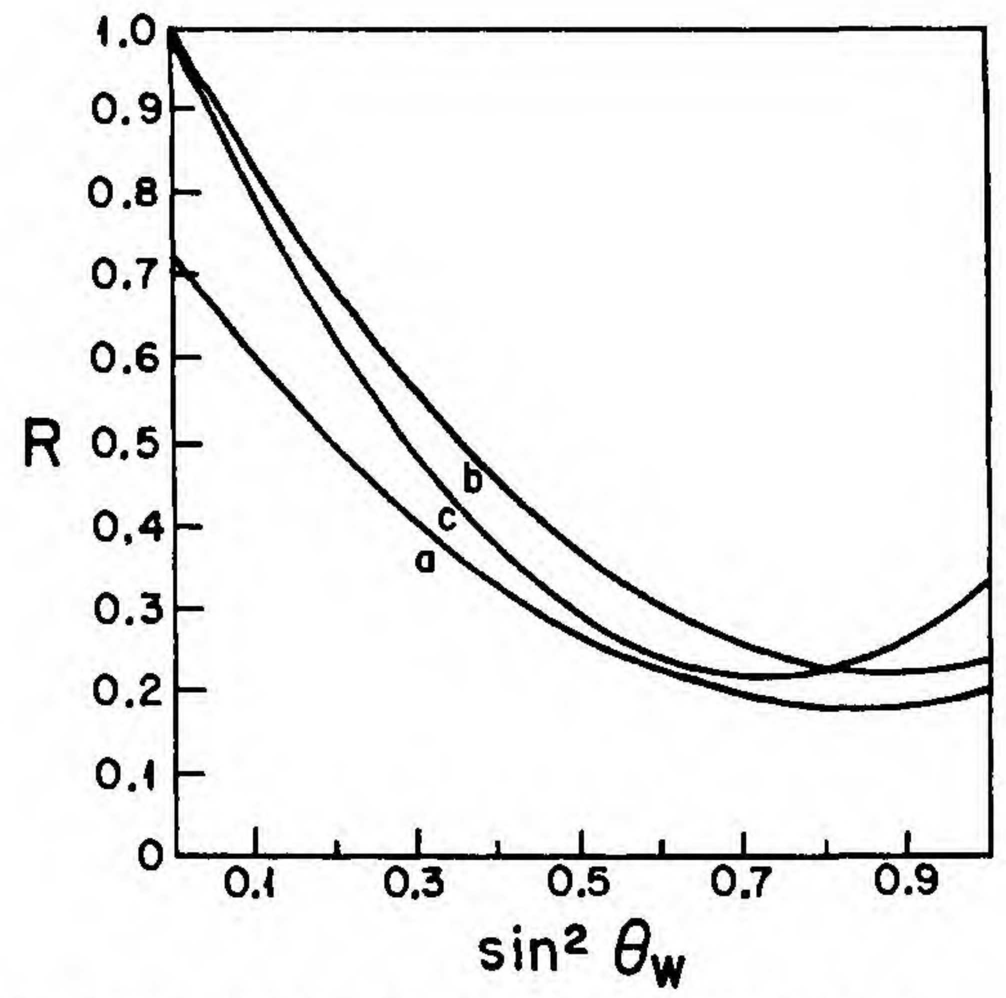


FIG. 1. Ratio R of Eq. (2) vs Weinberg angle θ_W. Curve a—resonant and nonresonant multipoles; curve b—resonant multipoles only; curve c—resonant multipoles in Lee's static-model calculation.

I wish to thank N. Christ, B. W. Lee, E. Paschos, and S. B. Treiman for conversations, and to acknowledge the hospitality of the Aspen Center for Physics, where this work was completed.

---

*Operated by Universities Research Association, Inc., under contract with the U. S. Atomic Energy Commission.

[1]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967). By the "Weinberg theory" we will always mean Weinberg's original model. There are, of course, many variations on the gauge-theory idea in which weak neutral currents are not present.

[2]H. S. Gurr, F. Reines, and H. W. Sobel, Phys. Rev. Lett. 28, 1406 (1972).

[3]B. W. Lee, Phys. Lett. 40B, 420 (1972).

[4]W. Lee, Phys. Lett. 40B, 423 (1972); CERN Gargamelle experiment, as reported by P. Musset, Bull. Am. Phys. Soc. 18, 73 (1973).

[5]D. H. Perkins, in *Proceedings of the XVI International Conference on High Energy Physics, Chicago-Batavia, Ill., 1972,* edited by J. D. Jackson and A. Roberts (NAL, Batavia, Ill., 1973), Vol. 4, p. 189.

[6]See, for example, footnote 18 of E. A. Paschos and L. Wolfenstein, Phys. Rev. D 7, 91 (1973).

[7]S. L. Adler, Ann. Phys. (N.Y.) 50, 189 (1968).

[8]G. F. Chew, F. E. Low, M. L. Goldberger, and Y. Nambu, Phys. Rev. 106, 1345 (1957).

[9]In Ref. 7 we also evaluate dispersive corrections to the nonresonant partial waves arising from (3, 3)-resonance exchange in the u channel. We omit these corrections in the present note, because they are small but costly to evaluate in terms of computer time.

[10]J. Campbell *et al.*, Phys. Rev. Lett. 30, 335 (1973); P. Schreiner and F. von Hippel, *ibid.* 30, 339 (1973).

[11]Curve c is obtained by setting the parameter η in Ref. 3 equal to zero; this corresponds to the static limit of the model of Ref. 7.

[12]F. J. Hasert *et al.*, Phys. Lett. 46B, 138 (1973).

PHYSICAL REVIEW D VOLUME 9, NUMBER 7 1 APRIL 1974

Nuclear charge-exchange corrections to leptonic pion production in the (3, 3)-resonance region

Stephen L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540 and National Accelerator Laboratory, Batavia, Illinois 60510*

Shmuel Nussinov† and E. A. Paschos
National Accelerator Laboratory, Batavia, Illinois 60510

(Received 7 December 1973)

We discuss nuclear charge-exchange corrections to leptonic pion production in the region of the (3, 3) resonance, both from a phenomenological viewpoint and from the evaluation of a detailed model for pion multiple scattering in the target nucleus. Using our analysis, we estimate the nuclear corrections needed to extract the ratio
$R = [\sigma(\nu_\mu + n \to \nu_\mu + n + \pi^0) + \sigma(\nu_\mu + p \to \nu_\mu + p + \pi^0)] / 2\sigma(\nu_\mu + n \to \mu^- + p + \pi^0)$
from neutral-current search experiments using ${}_{13}\mathrm{Al}^{27}$ and other nuclei as targets.

I. INTRODUCTION

Although weak-interaction experiments on hydrogen and deuterium targets are most readily interpreted theoretically, experimental considerations necessitate the use of complex nuclear targets in many of the current generation of accelerator neutrino experiments. As a result, in such experiments, corrections for nuclear effects must be made in order to extract free nucleon cross sections from the experimental data. Our aim in the present paper is to analyze these corrections in a particularly simple case: that of leptonic single-pion production in the region of the (3, 3) resonance. This reaction has gained prominence recently because measurement of the ratio

$$R = \frac{\sigma(\nu_\mu + n \to \nu_\mu + n + \pi^0) + \sigma(\nu_\mu + p \to \nu_\mu + p + \pi^0)}{2\sigma(\nu_\mu + n \to \mu^- + p + \pi^0)} \tag{1}$$

appears to be one of the better ways of searching for hadronic weak neutral currents.[1] Experiments measuring R use aluminum (and in some cases also carbon) as target materials, so the experimentally measured quantity is actually (T', T'' denote unobserved final target states)

$$R'(T) = \frac{\sigma(\nu_\mu + T \to \nu_\mu + T' + \pi^0)}{2\sigma(\nu_\mu + T \to \mu^- + T'' + \pi^0)}, \quad T = {}_{13}\mathrm{Al}^{27},\ {}_6\mathrm{C}^{12}\ . \tag{2}$$

As Perkins has emphasized,[2] nuclear charge-exchange effects can cause substantial differences between R and R', which makes reliable estimation of these effects an important ingredient in correctly interpreting the experiments. Fortunately, pion production in the (3, 3) region is also a particularly favorable case for theoretical analysis, primarily because the elasticity of the (3, 3) resonance implies that nuclear effects will not bring multipion or other more complex hadronic channels into play.

Our discussion is organized as follows. In Sec. II we introduce our basic phenomenological assumption: that leptonic pion production on a nuclear target may be represented as a two-step compound process in which pions are first produced from constituent nucleons with the free lepton-nucleon cross section (apart from a Pauli-principle reduction factor), and subsequently undergo a nuclear interaction independent of the identity of the leptons involved in the production step. This assumption allows us to isolate nuclear effects (pion charge exchange and absorption) in a 3×3 "charge-exchange matrix." We analyze the structure of the charge-exchange matrix in the particularly simple case of an isotopically neutral target nucleus. [For ${}_{13}\mathrm{Al}^{27}$, with a neutron excess of 1 and a corresponding isospin of $\frac{1}{2}$, the approximation of isotopic neutrality should be quite adequate. For ${}_6\mathrm{C}^{12}$, of course, no approximation is involved.] Our main phenomenological result is that parameters of the charge-exchange matrix, which can be measured in high-rate pion electroproduction experiments, can be used to calculate the nuclear corrections to the weak pion production process and in particular to give the connection between R' and R. In Sec. III we develop a detailed multiple-scattering model for the charge-exchange matrix. Our model is quite similar to a successful calculation by Sternheim and Silbar[3] of pion production in the (3, 3)-resonance region induced by protons incident on nuclear targets, and uses the nuclear pion absorption cross section which they determine (as well as the experimental pion-nucleon charge-exchange cross section) as

input. The principal difference between our calculation and that of Sternheim and Silbar (apart from obvious changes stemming from the fact that they deal with a strongly absorbed rather than a weakly interacting projectile) is that we use an improved approximation to the multiple-scattering problem, based on a one-dimensional scattering solution introduced by Fermi in the early days of neutron physics.[4] Using our model for the charge-exchange matrix, and a theoretical calculation of electroproduction and weak production of pions from free nucleons which has been described elsewhere,[5] we present detailed predictions for R' in the Weinberg weak-interaction theory and some of its variants. We also use the production model to test averaging approximations implicit in the phenomenological discussion of Sec. II. In Sec. IV we summarize briefly our conclusions. Three appendixes are devoted to mathematical details. In Appendix A we formulate and solve the one-dimensional scattering problem which forms the basis for the approximate solution of the three-dimensional multiple-scattering problem actually encountered in our model. To calibrate this approximation, in Appendix B we compare the approximate solution with the exact multiple-scattering solution for the simple case of isotropic scattering centers uniformly distributed within a sphere. Finally, in Appendix C we collect miscellaneous formulas for cross sections and for Pauli exclusion factors which are needed in the text.

II. PHENOMENOLOGY

A. Kinematics[6]

We consider the leptonic pion-production reaction

$$l(k_1) + T \rightarrow l'(k_2) + T' + \pi^{(\pm 0)}, \tag{3}$$

with k_1 and k_2 respectively the four-momenta of the initial and final lepton l and l', with T a nuclear target initially at rest in the laboratory, and with T' an unobserved final nuclear state. Let $k^2 = (k_1 - k_2)^2$ be the leptonic invariant four-momentum transfer squared, and let $k_0^L \equiv k_{10}^L - k_{20}^L$ be the laboratory leptonic energy transfer to the hadrons. Corresponding to the three pionic charge states in Eq. (3) we have three doubly differential cross sections with respect to the variables k^2 and k_0^L, which we denote by

$$\frac{d^2\sigma(ll'T;\pm 0)}{dk^2 dk_0^L}. \tag{4}$$

When the target T is a single nucleon N (of mass M_N), below the two-pion production threshold the recoil target T' must also be a single nucleon, and we can specify the kinematics more precisely. We write in this case

$$l(k_1) + N(p_1) \rightarrow l'(k_2) + N'(p_2) + \pi^{(\pm 0)}(q), \tag{5}$$

with the hadron four-momenta indicated in parentheses. We denote the final pion-nucleon isobar mass by W,

$$W^2 = (p_2 + q)^2. \tag{6a}$$

This variable is evidently related to the leptonic energy transfer k_0^L by

$$W^2 = M_N{}^2 - |k^2| + 2M_N k_0^L. \tag{6b}$$

B. Factorization assumption

We now introduce a factorization assumption which is basic to all of our subsequent arguments. We assume that leptonic pion production on a nuclear target may be regarded as a two-step compound process. In the first step of this process pions are produced from constituent nucleons of the target nucleus with the free lepton-nucleon cross section. In the second step the produced pions undergo a nuclear interaction, dependent on properties of the target nucleus and on the kinematic variables k_0^L and (possibly) k^2, but independent of the identities of the leptons involved in the first step. Since we are considering only excitation energies below the two-pion production threshold, the nuclear interaction in the second step involves only two types of processes, (i) scattering of the pion and (ii) absorption of the pion in two-nucleon or more complex nuclear processes. In particular, the two-pion production channel cannot come into play, and hence the factorization assumption allows us to write a simple matrix relation between the cross sections for leptonic pion production on nuclear and on free nucleon targets. We have

$$\begin{bmatrix} \dfrac{d^2\sigma(ll'\,{}_zT^A;+)}{dk^2dk_0^L} \\[2ex] \dfrac{d^2\sigma(ll'\,{}_zT^A;0)}{dk^2dk_0^L} \\[2ex] \dfrac{d^2\sigma(ll'\,{}_zT^A;-)}{dk^2dk_0^L} \end{bmatrix} = [M({}_zT^A;k^2k_0^L)] \begin{bmatrix} \dfrac{d^2\sigma(ll'N_T;+)_F}{dk^2dk_0^L} \\[2ex] \dfrac{d^2\sigma(ll'N_T;0)_F}{dk^2dk_0^L} \\[2ex] \dfrac{d^2\sigma(ll'N_T;-)_F}{dk^2dk_0^L} \end{bmatrix}, \tag{7}$$

with

$$\frac{d^2\sigma(ll'N_T;\pm 0)_F}{dk^2dk_0^L} = Z\frac{d^2\sigma(ll'p;\pm 0)_F}{dk^2dk_0^L} + (A-Z)\frac{d^2\sigma(ll'n;\pm 0)_F}{dk^2dk_0^L} \tag{8}$$

an appropriately weighted linear combination of free proton and free neutron cross sections. The subscript F indicates that these cross sections are to be averaged over the Fermi motion of the individual target nucleon in the nucleus, which substantially alters the shape (but not the integrated area) of the (3, 3) resonance[7] when plotted versus the excitation energy k_0^L. In writing Eq. (7) we have obviously used rotational symmetry, which implies that when the angular variables of the pion emerging from the nucleus are integrated over, no dependence remains on the angles characterizing the initial production of the pion. The matrix M appearing in Eq. (7) is a 3×3 "charge-exchange matrix" which is independent of the nature of the initial and final leptons l and l'. In addition to including pion scattering and absorption effects, M also takes into account the reduction of leptonic pion production in a nucleus resulting from the Pauli exclusion principle.[8] We will keep this effect in M in the ensuing phenomenological discussion, but when we make our multiple-scattering model in Sec. III we will separate it off as an explicit multiplicative factor.

C. Structure of M

Up to this point Eq. (7) applies to all nuclei, even those with a large neutron excess. In order to simplify the subsequent discussion we now restrict ourselves to the case of isotopically neutral targets, with the neutron excess and isotopic spin equal to zero.[9] (See Added Note preceding Acknowledgments.) As noted in the Introduction, this approximation is reasonable for the targets of greatest experimental interest. With this restriction, the pion charge structure of the matrix M_{fi} ($f, i = \pm, 0$) is that of the inclusive reaction

$$\pi_i + T(I=0) \to \pi_f + T'(\text{unobserved}), \tag{9a}$$

or equivalently, of the forward scattering process

$$\pi_i + \bar{\pi}_f + T(I=0) \to \pi_i + \bar{\pi}_f + T(I=0). \tag{9b}$$

Since the isospin of the system $\pi_i + \bar{\pi}_f$ can be either 0, 1, or 2, we conclude that the matrix M_{fi} involves three independent parameters. We introduce them by writing

$$M_{fi} = A(c-d)\psi_f\cdot\psi_i\,\psi_f^*\cdot\psi_i^* + Ad\,\psi_i\cdot\psi_i^*\psi_f\cdot\psi_f^* + A(1-c-2d)\psi_f\cdot\psi_i^*\psi_i\cdot\psi_f^*, \tag{10}$$

with ψ_i and ψ_f the isospin wave functions of π_i and π_f, respectively. Substituting

$$\psi_0 = \begin{pmatrix}0\\0\\1\end{pmatrix}, \quad \psi_\pm = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \pm i\\ 0\end{pmatrix} \tag{11}$$

and writing our Eq. (10) component by component, we get the basic form

$$\begin{pmatrix} M_{++} & M_{+0} & M_{+-}\\ M_{0+} & M_{00} & M_{0-}\\ M_{-+} & M_{-0} & M_{--}\end{pmatrix} = A\begin{pmatrix} 1-c-d & d & c\\ d & 1-2d & d\\ c & d & 1-c-d\end{pmatrix}. \tag{12}$$

It is useful to consider the form which the above equations take when no distinction is made between π^+ and π^-, but only between charged and neutral pions. Equation (7) is then replaced by

$$\begin{bmatrix} \dfrac{d^2\sigma(ll'{}_ZT^A;+)}{dk^2dk_0^L} + \dfrac{d^2\sigma(ll'{}_ZT^A;-)}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'{}_ZT^A;0)}{dk^2dk_0^L}\end{bmatrix} = [N({}_ZT^A;k^2k_0^L)]\begin{bmatrix} \dfrac{d^2\sigma(ll'N_T;+)_F}{dk^2dk_0^L} + \dfrac{d^2\sigma(ll'N_T;-)_F}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'N_T;0)_F}{dk^2dk_0^L}\end{bmatrix}, \tag{13}$$

with N a 2×2 matrix. When the target T is isotopically neutral, the matrix N can be expressed in terms of the parameters of Eq. (12), giving

$$\begin{pmatrix} N_{\text{ch ch}} & N_{\text{ch}\,0}\\ N_{0\,\text{ch}} & N_{00}\end{pmatrix} = A\begin{pmatrix} 1-d & 2d\\ d & 1-2d\end{pmatrix}. \tag{14}$$

The dependence on the parameter c has dropped out, leaving only two parameters which determine the nuclear corrections in this case.

D. Applications

Equations (13) and (14) can be applied in two ways. First, they can be used to generate a specific theoretical prediction for the ratio R' of Eq.

(2), by integrating with respect to k^2 and k_0^L [with the latter integration extending only over the (3, 3)-resonance region] to give

$$\sigma(\nu_\mu+T\to\nu_\mu+T'+\pi^0)=\int dk^2dk_0^L A(T;k^2k_0^L)\left\{d(T;k^2k_0^L)\left[\frac{d^2\sigma(\nu_\mu\nu_\mu N_T;+)_F}{dk^2dk_0^L}+\frac{d^2\sigma(\nu_\mu\nu_\mu N_T;-)_F}{dk^2dk_0^L}\right]\right.$$
$$\left.+[1-2d(T;k^2k_0^L)]\frac{d^2\sigma(\nu_\mu\nu_\mu N_T;0)_F}{dk^2dk_0^L}\right\},$$

$$\sigma(\nu_\mu+T\to\mu^-+T'+\pi^0)=\int dk^2dk_0^L A(T;k^2k_0^L)\left\{d(T;k^2k_0^L)\left[\frac{d^2\sigma(\nu_\mu\mu^-N_T;+)_F}{dk^2dk_0^L}+\frac{d^2\sigma(\nu_\mu\mu^-N_T;-)_F}{dk^2dk_0^L}\right]\right.$$
$$\left.+[1-2d(T;k^2k_0^L)]\frac{d^2\sigma(\nu_\mu\mu^-N_T;0)_F}{dk^2dk_0^L}\right\}. \tag{15}$$

In Sec. III we will evaluate Eq. (15) (and thus R') using our multiple-scattering model for M and the weak pion production calculation of Ref. 5 as inputs.[10]

The second way of applying Eqs. (13) and (14) is to use them in a purely empirical fashion to extract the charge-exchange matrix parameters $A(T;k^2k_0^L)$ and $d(T;k^2k_0^L)$ from a comparison of pion electroproduction on free nucleons with pion electroproduction on a nuclear target T. Specifically, we find from Eqs. (13) and (14) that

$$d(T;k^2k_0^L)=\frac{r(eeT)-r(eeN_T)}{[2-r(eeN_T)][1+r(eeT)]}, \tag{16}$$

with

$$r(eeT)=\left[\frac{d^2\sigma(eeT;+)}{dk^2dk_0^L}+\frac{d^2\sigma(eeT;-)}{dk^2dk_0^L}\right]\left[\frac{d^2\sigma(eeT;0)}{dk^2dk_0^L}\right]^{-1},$$
$$r(eeN_T)=\left[\frac{d^2\sigma(eeN_T;+)_F}{dk^2dk_0^L}+\frac{d^2\sigma(eeN_T;-)_F}{dk^2dk_0^L}\right]\left[\frac{d^2\sigma(eeN_T;0)_F}{dk^2dk_0^L}\right]^{-1}, \tag{17}$$

the electroproduction charged-pion-to-neutral-pion ratios on targets T and N_T, and

$$A(T;k^2k_0^L)=\left[\frac{d^2\sigma(eeT;+)}{dk^2dk_0^L}+\frac{d^2\sigma(eeT;0)}{dk^2dk_0^L}+\frac{d^2\sigma(eeT;-)}{dk^2dk_0^L}\right]\left[\frac{d^2\sigma(eeN_T;+)_F}{dk^2dk_0^L}+\frac{d^2\sigma(eeN_T;0)_F}{dk^2dk_0^L}+\frac{d^2\sigma(eeN_T;-)_F}{dk^2dk_0^L}\right]^{-1}. \tag{18}$$

Once $d(T;k^2k_0^L)$ and $A(T;k^2k_0^L)$ have been extracted from electroproduction data by use of Eqs. (16)–(18), they can be substituted into Eqs. (13) and (14) and used to calculate the nuclear corrections to *weak* pion production on the same target T. Note that Eqs. (16) and (17) for d are independent of the absolute normalization of the electron cross sections, and that A, which does depend on absolute normalization, appears as a simple multiplicative factor in both numerator and denominator of Eq. (2) for R'. Hence the relation between R and R' given by our empirical procedure is also independent of the absolute normalization of the electron cross sections used to extract A and d.

In many applications it is convenient to deal not with the doubly differential cross sections of Eq. (4), but rather with these cross sections integrated in excitation energy over the (3, 3)-resonance region,

$$\frac{d^2\sigma(ll'T;\pm 0)}{dk^2}=\int_{(3,3)\text{-resonance region}} dk_0^L\,\frac{d^2\sigma(ll'T;\pm 0)}{dk^2dk_0^L}. \tag{19}$$

In order to write simple formulas directly in terms of these integrated cross sections we note that, to a good first approximation, the k_0^L dependence of the doubly differential cross sections is governed by the dominant (3, 3) channel, and hence is independent of the identities of l and l' and of the pionic charge. This near-identity of excitation energy dependence allows us to make the averaging approximation of replacing Eqs. (7), (12), (13), and (14) by equations of identical form written directly in terms of the cross sections of Eq. (19),

$$\begin{bmatrix} \frac{d\sigma(ll'T;+)}{dk^2} \\ \frac{d\sigma(ll'T;0)}{dk^2} \\ \frac{d\sigma(ll'T;-)}{dk^2} \end{bmatrix} = [\overline{M}(T;k^2)] \begin{bmatrix} \frac{d\sigma(ll'N_T;+)}{dk^2} \\ \frac{d\sigma(ll'N_T;0)}{dk^2} \\ \frac{d\sigma(ll'N_T;-)}{dk^2} \end{bmatrix},$$

$$\begin{bmatrix} \frac{d\sigma(ll'T;+)}{dk^2} + \frac{d\sigma(ll'T;-)}{dk^2} \\ \frac{d\sigma(ll'T;0)}{dk^2} \end{bmatrix} = [\overline{N}(T,k^2)] \begin{bmatrix} \frac{d\sigma(ll'N_T;+)}{dk^2} + \frac{d\sigma(ll'N_T;-)}{dk^2} \\ \frac{d\sigma(ll'N_T;0)}{dk^2} \end{bmatrix}, \tag{20}$$

with $\overline{M}$ and $\overline{N}$ the lepton-independent matrices

$$[\overline{M}(T;k^2)] = \overline{A}\begin{pmatrix} 1-\overline{c}-\overline{d} & \overline{d} & \overline{c} \\ \overline{d} & 1-2\overline{d} & \overline{d} \\ \overline{c} & \overline{d} & 1-\overline{c}-\overline{d} \end{pmatrix},$$
$$[\overline{N}(T,k^2)] = \overline{A}\begin{pmatrix} 1-\overline{d} & 2\overline{d} \\ \overline{d} & 1-2\overline{d} \end{pmatrix}. \tag{21}$$

Because the excitation energy k_0^L has been integrated over,[7] we can use free nucleon cross sections on nucleon targets *at rest* in the right-hand side of Eq. (20); hence we have omitted the subscript F which indicated smearing of the production cross section over nucleon Fermi motion. (Any residual effects of nucleon Fermi motion on the excitation-energy-integrated pion production cross sections will, in this formulation, be absorbed in the phenomenological matrices $\overline{M}$ and $\overline{N}$.) The formulas for extracting $\overline{A}(T;k^2)$ and $\overline{d}(T;k^2)$ from electroproduction data are identical in form to Eqs. (16)–(18):

$$\overline{d}(T;k^2) = \frac{\overline{r}(eeT)-\overline{r}(eeN_T)}{[2-\overline{r}(eeN_T)][1+\overline{r}(eeT)]},$$
$$\overline{r}(eeT) = \left[\frac{d\sigma(eeT;+)}{dk^2}+\frac{d\sigma(eeT;-)}{dk^2}\right]\left[\frac{d\sigma(eeT;0)}{dk^2}\right]^{-1}, \tag{22a}$$
$$\overline{r}(eeN_T) = \left[\frac{d\sigma(eeN_T;+)}{dk^2}+\frac{d\sigma(eeN_T;-)}{dk^2}\right]\left[\frac{d\sigma(eeN_T;0)}{dk^2}\right]^{-1};$$

$$\overline{A}(T;k^2) = \left[\frac{d\sigma(eeT;+)}{dk^2}+\frac{d\sigma(eeT;0)}{dk^2}+\frac{d\sigma(eeT;-)}{dk^2}\right]\left[\frac{d\sigma(eeN_T;+)}{dk^2}+\frac{d\sigma(eeN_T;0)}{dk^2}+\frac{d\sigma(eeN_T;-)}{dk^2}\right]^{-1}. \tag{22b}$$

In terms of $\overline{d}$ and $\overline{A}$, the expression for R' analogous to Eqs. (2) and (15) is

$$R'(T) = \frac{\int dk^2 \overline{A}(T;k^2)\left\{\overline{d}(T;k^2)\left[\frac{d\sigma(\nu_\mu\nu_\mu N_T;+)}{dk^2}+\frac{d\sigma(\nu_\mu\nu_\mu N_T;-)}{dk^2}\right]+[1-2\overline{d}(T;k^2)]\frac{d\sigma(\nu_\mu\nu_\mu N_T;0)}{dk^2}\right\}}{2\int dk^2 \overline{A}(T;k^2)\left\{\overline{d}(T;k^2)\left[\frac{d\sigma(\nu_\mu\mu^- N_T;+)}{dk^2}+\frac{d\sigma(\nu_\mu\mu^- N_T;-)}{dk^2}\right]+[1-2\overline{d}(T;k^2)]\frac{d\sigma(\nu_\mu\mu^- N_T;0)}{dk^2}\right\}}. \tag{23}$$

Equations (20)–(23) are in a form convenient for direct comparison with experimental data, and constitute our principal phenomenological result.

We continue by introducing one further averaging approximation. To the extent that $\overline{d}(T;k^2)$ and $\overline{A}(T;k^2)$ are slowly varying functions of k^2 (and this is suggested by the numerical work of Sec. III) we can replace them by average values $\tilde{d}(T)$ and $\tilde{A}(T)$ in the integrals of Eq. (23). The parameter $\tilde{A}(T)$ then cancels between numerator and denominator and the integration over k^2 can be explicitly carried out. We are left with a simple formula relating R' to R,

$$R'(T) = R\frac{\tilde{d}(T)\tilde{r}(\nu_\mu\nu_\mu N_T)+1-2\tilde{d}(T)}{\tilde{d}(T)\tilde{r}(\nu_\mu\mu^- N_T)+1-2\tilde{d}(T)}, \tag{24}$$

with

$$\tilde{r}(\nu_\mu\nu_\mu N_T) = \frac{\sigma(\nu_\mu\nu_\mu N_T;+)+\sigma(\nu_\mu\nu_\mu N_T;-)}{\sigma(\nu_\mu\nu_\mu N_T;0)},$$
$$\tilde{r}(\nu\mu^- N_T) = \frac{\sigma(\nu_\mu\mu^- N_T;+)+\sigma(\nu_\mu\mu^- N_T;-)}{\sigma(\nu_\mu\mu^- N_T;0)}, \tag{25}$$

the charged-pion-to-neutral-pion ratios produced on an average nucleon target by neutral and charged weak currents, respectively. In the approximation of Eq. (24), nuclear charge-exchange effects are isolated in the single parameter $\tilde{d}(T)$. This description is particularly useful for giving a simple comparison of the charge-exchange corrections expected for different nuclear targets T.

E. Discussion

We conclude by pointing out an experimental problem which will limit the direct applicability of the phenomenological results of Eqs. (16)–(23). In all of the above equations, we have assumed that the angular variables of the produced pion are unobserved, which corresponds to an experimental situation in which the acceptance for produced pions is 4π sr. However, in realistic experiments observing the weak production and electroproduction of pions, the pion acceptance will, in general, be rather small. Since the pion angular distributions *do* depend on the leptons involved in the production process,[11] the introduction of acceptance restrictions will tend to spoil the simple relation between nuclear charge-exchange corrections to weak production and electroproduction of pions which we have developed above. There are two possible ways of dealing with this problem. One would be to simply go ahead and apply Eqs. (21)–(23) to the limited-acceptance case, interpreting the cross sections on T and N_T as being limited to the actual pion acceptance. If both the value of $\bar{d}$ extracted from electroproduction[12] and the pion charge ratios observed in weak production were found to be only weakly acceptance-dependent, one would have an *a posteriori* justification for applying the phenomenological recipe of Eq. (23) to the acceptance-limited case. An alternative procedure would be to develop a detailed model for the charge-exchange parameters d, c, and A, and then to numerically fold these charge-exchange corrections into experimental or theoretical cross sections for pion production on a free-nucleon target, taking acceptance limitations into account. Although, in this approach, one would forego the possibility of direct phenomenological application of electroproduction data, a comparison of the theory with electroproduction experiments on nuclear targets would still be essential to test (and possibly revise) the charge-exchange model. Once validated in this way, the charge-exchange parameters could be substituted into Eqs. (15) and (23) to generate predictions for weak-production experiments. The question of constructing a suitable model for the charge-exchange parameters will be pursued further in Sec. III.

III. MULTIPLE-SCATTERING MODEL

We proceed in this section to develop a detailed multiple-scattering model for nuclear charge-exchange corrections. Our motivations are, first, to get an estimate of the magnitude of charge-exchange corrections to be expected for various target nuclei, and second, as discussed above, to facilitate comparison with experiment in the realistic case in which there are pion acceptance limitations.

A. Formulation of the model

Our model closely resembles (with differences which we explain below) a successful semiclassical treatment of $\pi^{\pm}$ production in proton-nucleus collisions which has been given by Sternheim and Silbar.[3] The ingredients of the model are as follows:

(1) We regard the target nucleus as a collection of independent nucleons, distributed spatially according to the density profile determined by electron scattering experiments. For aluminum and lighter nuclei, it is convenient to parameterize the nucleon density in the so-called "harmonic well" form

$$\rho(r)=\rho(0)e^{-r^2/R^2}\left[1+c\frac{r^2}{R^2}+c_1\left(\frac{r^2}{R^2}\right)^2\right], \tag{26}$$

with the values of the various parameters given in Table I.

(2) In discussing pion multiple scattering within the target nucleus, we regard the nucleons as fixed within the nucleus, thus neglecting Fermi motion and nucleon recoil effects. [A numerical estimate of the importance of these effects will be made in Sec. III B 2 below.] This approximation allows us to characterize interactions of the pion with the constituent nucleons by a unique center-of-mass energy W, related to the lepton energy transfer k_0^L by Eq. (6b). Through all stages of the multiple scattering we approximate the target nucleus to be isotopically neutral, composed of equal numbers of protons and neutrons. (See Added Note.)

(3) Interactions of pions in the nucleus are

TABLE I. Nuclear density parameters.[a,b]

Nucleus $_ZT^A$	c	c_1	R (F) (Ref. c)	$R\rho(0)$ (F)$^{-2}$
$_5B^{10}$	1	0	2.45	0.251
$_6C^{12}$	1.333	0	2.41	0.268
$_7N^{14}$	1.667	0	2.46	0.263
$_8O^{16}$	1.600	0	2.75	0.247
$_{13}Al^{27}$	2.000	0.667	1.76	0.241

[a]The data are taken from H. R. Collard, L. R. B. Elton, and R. Hofstadter, in Landolt-Börnstein: *Numerical Data and Functional Relationships; Nuclear Radii*, edited by K.-H. Hellwege (Springer, Berlin, 1967), New Series, Group I, Vol. 2.

[b]The density $\rho(r)$ is normalized so that $\int d^3r\rho(r)=A$.

[c]For the first four nuclei, the rms charge radius is equal to R. For aluminum, the rms charge radius corresponding to the listed parameters is 2.91 F.

treated in the approximation of complete incoherence, involving the use of pion-nucleon cross sections rather than scattering amplitudes in the multiple-scattering calculation. In the region of the (3, 3) resonance, pion production and more complex hadron production channels are closed, and so there are only two relevant cross sections. The first is the cross section per nucleon $\sigma_{\rm abs}(W)$ for pion absorption via various nuclear processes; for this quantity we use the best-fit value obtained by Sternheim and Silbar in their study of pion production by protons,

$$\sigma_{\rm abs}(W)=\begin{cases}0, & T_\pi<0.788M_\pi\\ 22\ \text{mb}\times\dfrac{T_\pi-0.788M_\pi}{2.077M_\pi}, & T_\pi>0.788M_\pi\end{cases} \tag{27}$$

where

$$T_\pi=\frac{W^2-(M_N+M_\pi)^2}{2M_N}\ .$$

To allow for the considerable uncertainties in this expression for $\sigma_{\rm abs}$, we examine numerically the effect on the charge-exchange corrections of multiplying Eq. (27) by factors of $\frac{1}{2}$ or 2. (See also Added Note.) The second cross section needed is the usual elastic cross section for pion-nucleon scattering. Since in the (3, 3) region the $I=\frac{1}{2}$ pion-nucleon cross section is very small, we neglect it entirely and regard all pion-nucleon scattering as proceeding through the $I=\frac{3}{2}$ channel. The elastic cross section is then simply proportional to the cross section

$$\sigma_{\pi^+p}(W)\ , \tag{28}$$

for which a simple parameterization is given in Appendix C. In order to solve the pion multiple-scattering problem, we actually need the differential cross section for elastic scattering; in the approximation of (3, 3) dominance, this is given by

$$\frac{d\sigma_{\rm elastic}}{d\Omega}\propto\sigma_{\pi^+p}(W)(1+3\cos^2\phi)\ , \tag{29}$$

with ϕ the pion scattering angle.

(4) When a pion is produced by leptons incident on a nucleus or undergoes subsequent rescatterings, with small momentum transfer to the nuclear system, the corresponding production or scattering cross section is reduced by the Pauli exclusion principle.[6] We take this effect into account, within the framework of the independent nucleon picture, by multiplying the pion-leptoproduction cross section and the pion-nucleon rescattering cross section by respective reduction factors $g(W,k^2)$ and $h(W,\phi)$. Formulas for these factors are given in Appendix C. Neutrino quasielastic scattering experiments at small momentum transfer k^2 provide some empirical evidence for the presence of the production factor g. The argument for including h is less compelling, since we are using a semiclassical picture, with fixed constituent nucleons, for treating the pion multiple scattering in the nuclear medium, and in a semiclassical picture there are no Pauli effects. To take this objection[13] into account, in the numerical work below we also calculate results for the case in which h is replaced by unity.

(5) The approximation of keeping only $I=\frac{3}{2}$ pion-nucleon scattering allows us to reduce the problem of calculating the charge-exchange matrix M to a one-component scattering problem. To see this we let

$$\psi_i=\begin{pmatrix}n_i(\pi^+)\\ n_i(\pi^0)\\ n_i(\pi^-)\end{pmatrix} \tag{30}$$

denote the pion charge multiplicities initially present in a beam of pions, at a fixed isobar energy W. A simple isospin Clebsch analysis then shows that when the pion beam undergoes a single scattering on an equal mixture of protons and neutrons through the $I=\frac{3}{2}$ channel, the effect is to replace ψ by $Q\psi$, with Q the matrix

$$Q=\begin{pmatrix}\frac{5}{6} & \frac{1}{6} & 0\\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6}\\ 0 & \frac{1}{6} & \frac{5}{6}\end{pmatrix}. \tag{31}$$

Obviously, the natural way to describe a multiple-scattering process in which Q acts on ψ repeatedly is to decompose ψ into a sum of eigenvectors of Q. These eigenvectors, with their corresponding eigenvalues λ, are

$$q_1=\begin{pmatrix}1\\1\\1\end{pmatrix},\quad \lambda_1=1$$

$$q_2=\begin{pmatrix}1\\0\\-1\end{pmatrix},\quad \lambda_2=\tfrac{5}{6} \tag{32}$$

$$q_3=\begin{pmatrix}1\\-2\\1\end{pmatrix},\quad \lambda_3=\tfrac{1}{2}$$

and the decomposition reads

$$\psi=\sum_{k=1}^{3}C_k q_k,$$
$$\begin{aligned}C_1&=\tfrac{1}{3}[n_i(\pi^+)+n_i(\pi^0)+n_i(\pi^-)],\\ C_2&=\tfrac{1}{2}[n_i(\pi^+)-n_i(\pi^-)],\\ C_3&=-\tfrac{1}{3}n_i(\pi^0)+\tfrac{1}{6}[n_i(\pi^+)+n_i(\pi^-)].\end{aligned}\tag{33}$$

The effect of a multiple-scattering process on Eq. (33) will be to lead to a final pion multiplicity state ψ_f, related to ψ_i by

$$\psi_f=\sum_{k=1}^{3}f(\lambda_k)C_k q_k,\tag{34}$$

with $f(\lambda)$ a function of the eigenvalue λ which contains all geometric and dynamical information concerning nuclear parameters, magnitudes of cross sections, etc. Taking now ψ_i to be the initial distribution of lepto-produced pions in target T,

$$\psi_i=g(W,k^2)\begin{bmatrix}\dfrac{d^2\sigma(ll'N_T;+)}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'N_T;0)}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'N_T;-)}{dk^2dk_0^L}\end{bmatrix}\tag{35}$$

and ψ_f to be the distribution of exiting pions,

$$\psi_f=\begin{bmatrix}\dfrac{d^2\sigma(ll'T;+)}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'T;0)}{dk^2dk_0^L}\\[2ex] \dfrac{d^2\sigma(ll'T;-)}{dk^2dk_0^L}\end{bmatrix},\tag{36}$$

we find that the connection of Eq. (34) takes the form of Eqs. (7) and (12), with

$$\begin{aligned}&A=g(W,k^2)a,\quad a=f(1)\\ &c=\tfrac{1}{3}-\tfrac{1}{2}f(\tfrac{5}{6})/f(1)+\tfrac{1}{6}f(\tfrac{1}{2})/f(1),\\ &d=\tfrac{1}{3}[1-f(\tfrac{1}{2})/f(1)].\end{aligned}\tag{37}$$

(6) We turn finally to the function $f(\lambda)$, which contains the dynamical details of pion multiple scattering in the nucleus. The precise statement of the problem defining $f(\lambda)$ is as follows: We introduce an initial distribution of monoenergetic pions into a nucleus, with the pion density proportional to the nuclear density [given by Eq. (26)]. The pions are multiply scattered, with absorption cross section given by Eq. (27) and with elastic scattering cross section given by Eq. (29). At each elastic scattering the pion number is multiplied by a factor λ. The function $f(\lambda)$ is then defined as the expected number of pions eventually emerging from the nuclear medium, normalized to unit integrated initial pion density.

To get a simple (and, it turns out, surprisingly accurate) approximation to $f(\lambda)$, we replace the actual angular distribution [Eq. (29) times $h(W,\cos\phi)$] by a modified elastic scattering distribution, in which all forward-hemisphere scattering ($0\le\phi\le\pi/2$) is projected onto the forward direction ($\phi=0$), and all backward-hemisphere scattering ($\pi/2\le\phi\le\pi$) is projected onto the backward direction ($\phi=\pi$). In this approximation, once a pion is produced in the nucleus, it scatters back and forth along its initial line of motion until it either is absorbed or it leaves the nucleus. Since both the initial pion distribution and the interaction probabilities are proportional to nucleon density, the nucleon density profile along the line can be scaled out of the problem by an appropriate change in length variable. Thus, for each line passing through the nucleus the expected fraction of pions which exit is independent of the density profile along the line, but depends only on the integrated density along the line (the so-called optical thickness), which we denote by L. Once we have solved for the one-dimensional exit fraction $f(\lambda,L)$, we need only average over the distribution of optical thickness in the nucleus to get an expression for $f(\lambda)$.

To put these remarks in quantitative form, let us take the central nucleon density $\rho(0)$ as the "standard density" relative to which densities elsewhere in the nucleus are measured. For given impact parameter b relative to the center of the nucleus, the optical thickness is then given by

$$\begin{aligned}L(b)&=\int_{-\infty}^{\infty}dz\,e^{-(z^2+b^2)/R^2}\left[1+c\left(\frac{z^2+b^2}{R^2}\right)+c_1\left(\frac{z^2+b^2}{R^2}\right)^2\right]\\ &=R\pi^{1/2}e^{-b^2/R^2}\left\{1+c\left(\frac{1}{2}+\frac{b^2}{R^2}\right)+c_1\left[\frac{3}{4}+\frac{b^2}{R^2}+\left(\frac{b^2}{R^2}\right)^2\right]\right\}.\end{aligned}\tag{38}$$

Averaging over impact parameters, the relation between $f(\lambda)$ and $f(\lambda, L)$ is given by

$$f(\lambda)=\frac{\int_0^\infty bdb\, L(b) f(\lambda, L(b))}{\int_0^\infty bdb\, L(b)}. \tag{39}$$

The one-dimensional problem defining $f(\lambda, L)$ is formulated precisely as follows: We consider a uniform one-dimensional medium of length L, in which pions are uniformly initially produced moving (say) to the right. The pions propagate in the medium with inverse interaction length κ, given in terms of the nucleon density and the absorption and scattering cross sections by

$$\begin{aligned}\kappa&=\rho(0)\sigma_{\text{tot}}\,,\\ \sigma_{\text{tot}}&=\sigma_{\text{abs}}(W)+\tfrac{1}{3}\sigma_{\pi^+p}(W)[h_+(W)+h_-(W)]\,.\end{aligned} \tag{40}$$

The factors h_+ and h_- describe the forward- and backward-hemisphere projections of the Pauli reduction factor $h(W, \phi)$,

$$\begin{aligned}h_+&=\tfrac{1}{2}\int_0^{\pi/2}\sin\phi\, d\phi(1+3\cos^2\phi)h(W,\phi)\,,\\ h_-&=\tfrac{1}{2}\int_{\pi/2}^{\pi}\sin\phi\, d\phi(1+3\cos^2\phi)h(W,\phi)\,,\end{aligned} \tag{41}$$

and are explicitly calculated in Appendix C. At each interaction the pions are forward-scattered with probability μ_+ and back-scattered with probability μ_- (and, of course, absorbed with probability $1-\mu_+-\mu_-$), with

$$\mu_\pm=\tfrac{1}{3}\sigma_{\pi^+p}(W)h_\pm(W)/\sigma_{\text{tot}}\,, \tag{42}$$

and, concomitantly with each scattering, the pion number is multiplied by a factor λ. The desired quantity $f(\lambda, L)$ is the expected number of pions eventually emerging from the medium, normalized to unit integrated initial pion density. An explicit expression for f is calculated in Appendix A [see Eq. (A12) and Eqs. (A25)–(A27)], as well as expressions for f_+ and f_-, the expected fractions of pions eventually emerging with and without a net reversal of direction of motion along the line. In Appendix B we compare the approximate solution for $f(\lambda)$ given by Eq. (39) with the exact solution in the simple geometry of a uniform sphere composed of material which scatters isotropically, and find very satisfactory agreement. Since the actual angular distribution of interest to us, $1+3\cos^2\phi$, is already peaked in the backward and forward directions,[14] our approximation should be at least as accurate for this case as it is for handling isotropic scattering.

This completes the specification of our multiple-scattering model. As we have already noted, it closely resembles the calculation of Sternheim and Silbar, and the reader is referred to Ref. 3 for an excellent, detailed analysis of the approximations and physical assumptions which are involved. The aspects in which our model differs from that of Ref. 3 are the following: (1) We take into account the diffuseness of the nuclear edge, rather than treating the nucleon distribution as a uniform sphere; (2) we take Pauli exclusion effects into account in a crude way; and (3) we use an improved approximation for solving the pion multiple-scattering problem. Instead of using the back-forward approximation described above, Sternheim and Silbar use the considerably less accurate approximation of treating all scattering as purely forward scattering. A comparison of their approximation with the exact solution, in the case of a uniform sphere composed of material which scatters isotropically, is given in Appendix B.

B. Numerical calculations

We turn now to numerical calculations, in which we combine our model for nuclear charge-exchange corrections with the theory of electroproduction and weak production of pions from free-nucleon targets developed in Ref. 5. For the hadronic weak neutral current, we adopt the Weinberg-model form[15]

$$J_\lambda^{\text{neutral}}=J_\lambda^{V3}+J_\lambda^{A3}-2\sin^2\theta_W J_\lambda^{\text{em}}\,; \tag{43}$$

we will say a few words below about variants of this model in which Eq. (43) contains an additional isoscalar current. We assume throughout an incident lab neutrino energy $k_{10}^L=1$ GeV and a nucleon elastic form factor[2]

$$g_A(k^2)=\frac{1.24}{[1+k^2/(0.9\text{ GeV}/c)^2]^2}\,, \tag{44}$$

and take integrations over the (3, 3)-resonance region to extend from the pion production threshold up to a maximum isobar mass of $W=1.47$ GeV. In our calculations on aluminum, we weight the free-nucleon production cross sections according to the actual neutron/proton ratio in aluminum (i.e., we take $N_T=13p+14n$), but as emphasized above, we adopt the approximation of isotopic neutrality in calculating charge-exchange corrections.

1. *Calculation of R′ from Eq. (15) (with Fermi motion neglected)*

In Table II we present results for the ratio R' on an aluminum target, calculated by using Eq. (15) to fold the W-dependent charge-exchange matrix into the production cross sections from a free-nucleon target at rest [i.e., we neglect the Fermi-motion average symbolized by the sub-

script F in Eq. (15)]. In the second column we tabulate

$$R(N_T)=\frac{\sigma(\nu_\mu+N_T\to\nu_\mu+N_T'+\pi^0)}{2\sigma(\nu_\mu+N_T\to\mu^-+N_T''+\pi^0)}, \tag{45}$$

which is the ratio predicted by the production model when no charge-exchange corrections are made. In the third through seventh columns we tabulate values of the charge-exchange-corrected ratio R' obtained under various alternative assumptions. The column labeled "no variations" is the result obtained from the multiple-scattering model of Sec. III A above; the next three columns show how this result changes when the Pauli factors h in Eqs. (40) and (42) are replaced by unity, or when the absorption cross section of Eq. (27) is modified. The predictions for R' are evidently quite insensitive to these variations. The seventh column gives the result for R' when all isoscalar multipoles are omitted. Since the isoscalar multipoles only contribute *quadratically* to R',[16] this column gives a lower bound on R' for any variant of the Weinberg theory in which the hadronic neutral current differs from Eq. (43) by purely isoscalar terms. In the final column we have used our production and charge-exchange calculations to generate simulated pion weak-production and electroproduction cross sections on aluminum, which are then used to evaluate the lower bound on R' derived by Albright *et al.*[17] in the isoscalar-target approximation,

$$\begin{aligned}
&R'({}_{13}\mathrm{Al}^{27})\geq\frac{1}{4}\left\{[\bar r(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27})-1]^{1/2}-2\sin^2\theta_W\left[\frac{V^0_{\mathrm{em}}}{\sigma(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27};0)}\right]^{1/2}\right\}^2,\\
&\bar r(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27})=\frac{\sigma(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27};+)+\sigma(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27};-)}{\sigma(\nu_\mu\mu^-{}_{13}\mathrm{Al}^{27};0)},\\
&V^0_{\mathrm{em}}=\frac{G^2\cos^2\theta_C}{\pi}\,\frac{1}{4\pi\alpha^2}\int(k^2)^2dk^2\,\frac{d\sigma(ee\,{}_{13}\mathrm{Al}^{27};0)}{dk^2}.
\end{aligned}\tag{46}$$

We see that the bound of Eq. (46) provides a satisfactory estimate of R' for small values of $\sin^2\theta_W$.

We turn next to Table III, where we have tabulated charged-pion to neutral-pion production ratios for the usual charged weak current. The first column gives the standard 5 : 1 prediction for an isotopically neutral target, assuming complete $I=\frac{3}{2}$ dominance. When $I=\frac{1}{2}$ multipoles are taken into account,[18] the prediction is lowered to 3.67 : 1, as shown in the second column. Finally, in the third column we give the prediction of 2.63 : 1 which results when Eq. (15) and its analog for charged pions are used to fold in charge-exchange corrections for aluminum.[19] It would obviously be very desirable to try to check this prediction for $\bar r$ simultaneously with the experimental determination of R'.

TABLE II. Calculations of $R'({}_{13}\mathrm{Al}^{27})$ based on Eq. (15).

$\sin^2\theta_W$	$R(N_T)$[a]	$R'({}_{13}\mathrm{Al}^{27})$ No variations	Pauli factors $h\to 1$	with $\frac{1}{2}\sigma_{\mathrm{abs}}$	with $2\sigma_{\mathrm{abs}}$	Isoscalar multipoles omitted	Simulated Albright *et al.* lower bound on $R'({}_{13}\mathrm{Al}^{27})$
0	0.697	0.422	0.396	0.411	0.435	0.422	0.408
0.1	0.573	0.346	0.325	0.337	0.356	0.346	0.321
0.2	0.465	0.280	0.264	0.273	0.289	0.280	0.245
0.3	0.374	0.225	0.212	0.220	0.232	0.225	0.179
0.4	0.300	0.180	0.170	0.176	0.186	0.180	0.123
0.5	0.242	0.146	0.138	0.143	0.150	0.145	0.078
0.6	0.200	0.122	0.115	0.119	0.125	0.120	0.043
0.7	0.175	0.108	0.102	0.106	0.110	0.106	0.019
0.8	0.166	0.104	0.099	0.102	0.107	0.102	0.004
0.9	0.174	0.111	0.106	0.109	0.113	0.108	0.000
1.0	0.198	0.128	0.123	0.126	0.131	0.125	0.006

[a]The numbers in this column are slightly smaller than those plotted in curve *a* of S. Adler [Phys. Rev. D 9, 229 (1974)], because we have reduced the axial-vector mass parameter [See Eq. (44)] from 1.0 to 0.9 GeV/c in the present calculation, and have also weighted the production cross sections according to the actual neutron/proton ratio in aluminum.

TABLE III. Charged-pion-to-neutral-pion ratio $\bar{r}(\nu_\mu \mu^- T)$.

$\bar{r}(\nu_\mu \mu^- n + p)$ pure (3, 3) approximation	$\bar{r}(\nu_\mu \mu^- N_T)$ with $I=\frac{1}{2}$ corrections	$\bar{r}'(\nu_\mu \mu^- {}_{13}\mathrm{Al}^{27})$ from Eq. (15) and charged-pion analog	$\bar{r}'(\nu_\mu \mu^- {}_{13}\mathrm{Al}^{27})$ from Eq. (48)
5	3.67	2.63	2.68

2. *Averaging approximations, comparison of different nuclei, and estimate of nucleon motion effects*

We conclude with a test of the averaging approximations introduced in Sec. II and a discussion of related topics. To study Eqs. (20)–(23), we fold together the electroproduction and charge-exchange models, as in Eq. (15), to give simulated data for pion electroproduction cross sections on aluminum. Substituting this data into Eqs. (21) and (22) then gives the values for $\bar{d}$ and $\bar{A}$ tabulated in Table IV. The charge-exchange parameters obtained this way are seen to be nearly independent of the incident electron energy k_{10}^L, and are slowly varying functions of k^2 except in the region $k^2 \leq 0.3$, where Pauli exclusion effects and $I=\frac{1}{2}$ multipoles arising from the pion exchange graph become important. Substituting the 2-GeV/c values of $\bar{d}$ and $\bar{A}$ into Eq. (23), and continuing to use our production model for the neutrino cross sections, gives the values of R' tabulated in the second column of Table V. In the third column we transcribe from Table II the values of R' obtained directly from Eq. (15); the good agreement indicates that the averaging approximation is working.

We turn next to the "double-averaged" approximation of Eqs. (24) and (25). We define the tilded charge-exchange parameters by averaging the charge-exchange matrix over the leading W-dependent part of the production cross section as obtained in the static approximation[20]:

$$\tilde{f}(\lambda) = \frac{\int dW q(W)^{-1} \sigma_{(3,3)}(W) f(\lambda)}{\int dW q(W)^{-1} \sigma_{(3,3)}(W)},$$
$$\tilde{a} = \tilde{f}(1),$$
$$\tilde{c} = \tfrac{1}{3} - \tfrac{1}{2}\tilde{f}(\tfrac{5}{3})/\tilde{f}(1) + \tfrac{1}{6}\tilde{f}(\tfrac{1}{2})/\tilde{f}(1),$$
$$\tilde{d} = \tfrac{1}{3}[1 - \tilde{f}(\tfrac{1}{2})/\tilde{f}(1)]. \tag{47}$$

Expressions for the resonant pion-nucleon scattering cross section $\sigma_{(3,3)}(W)$ and the pion momentum $q(W)$ are given in Appendix C. Evaluating Eq. (47) for aluminum gives $\tilde{d}({}_{13}\mathrm{Al}^{27}) = 0.162$, which, when substituted into Eq. (24) along with the charged-to-neutral ratios tabulated in the second and third columns of Table VI, gives the predictions for R' tabulated in the fourth column. These agree well with the corresponding values of R' obtained directly from Eq. (15). As another test of the "double averaged" approximation, we consider the formula giving the charge-exchange corrections to the charged-to-neutral ratio $\bar{r}$,

$$\bar{r}'(\nu_\mu \mu^- {}_{13}\mathrm{Al}^{27}) = \frac{2\tilde{d}({}_{13}\mathrm{Al}^{27}) + [1 - \tilde{d}({}_{13}\mathrm{Al}^{27})]\bar{r}(\nu_\mu \mu^- N_T)}{1 - 2\tilde{d}({}_{13}\mathrm{Al}^{27}) + \tilde{d}({}_{13}\mathrm{Al}^{27})\bar{r}(\nu_\mu \mu^- N_T)}. \tag{48}$$

Substituting $\bar{r} = 3.67$, $\tilde{d} = 0.162$ into Eq. (48) gives $\bar{r}' = 2.68$, as tabulated in the final column of Table III. This again is in close agreement with the value of $\bar{r}'$ obtained directly from Eq. (15).

As we remarked in Sec. II, the double-averaged approximation provides a convenient format for comparing charge-exchange effects in different

TABLE IV. Simulated $\bar{d}({}_{13}\mathrm{Al}^{27}; k^2)$ and $\bar{A}({}_{13}\mathrm{Al}^{27}; k^2)$ obtained from electroproduction and charge-exchange correction models.

	$k_{10}^L = 2$ GeV/c		$k_{10}^L = 6$ GeV/c	
k^2 (GeV/c)^2	$\bar{d}({}_{13}\mathrm{Al}^{27}; k^2)$	$\bar{A}({}_{13}\mathrm{Al}^{27}; k^2)$	$\bar{d}({}_{13}\mathrm{Al}^{27}; k^2)$	$\bar{A}({}_{13}\mathrm{Al}^{27}; k^2)$
0	0.191	0.606	0.188	0.608
0.1	0.181	0.688	0.179	0.682
0.2	0.169	0.702	0.167	0.694
0.3	0.164	0.702	0.162	0.694
0.4	0.160	0.698	0.158	0.690
0.6	0.157	0.694	0.154	0.684
0.8	0.156	0.690	0.152	0.680
1.0	0.155	0.690	0.150	0.676
1.4	0.155	0.688	0.147	0.672
1.8	0.156	0.686	0.145	0.666

TABLE V. Test of first averaging approximation for R' ($_{13}Al^{27}$).

$\sin^2\theta_W$	R' ($_{13}Al^{27}$) From Eq. (23)	From Eq. (15)
0	0.433	0.422
0.1	0.355	0.346
0.2	0.288	0.280
0.3	0.232	0.225
0.4	0.186	0.180
0.5	0.151	0.146
0.6	0.127	0.122
0.7	0.113	0.108
0.8	0.109	0.104
0.9	0.116	0.111
1.0	0.134	0.128

nuclei. In Table VII we have tabulated the charge-exchange parameters $\bar{a}$, $\bar{c}$, and $\bar{d}$ for a range of light and medium-weight nuclei up to aluminum. The key point to notice is that the parameter $\bar{d}$ is slowly varying, indicating that charge-exchange effects in different medium-weight targets, such as, for example, freon (CF_3Br) and aluminum, should be quite similar.

Finally, we apply the double-averaged approximation to estimate the effect on our numerical results of including nucleon Fermi motion and nucleon recoil. Obviously, to include nucleon motion in a realistic way one would have to go outside the framework of the one-speed scattering theory used above, since once the nucleons are not regarded as fixed the pion changes energy in each collision. Rather than attempting to follow these energy changes in detail (which would require an elaborate numerical calculation), we adopt a simple approximation which can be treated by the methods used above. We observe that in the (3, 3)-resonance region typical nucleon recoil momenta are of the same order as the nucleon Fermi momentum ($\sim 1.6\, M_\pi/c$); hence a rough estimate of nucleon-recoil and Fermi-motion effects should be given by the simple randomizing approximation of regarding the pion energy as a constant throughout its motion in the nucleus, but replacing the pion-production and charge-exchange-scattering cross sections by corresponding cross sections which are smeared over nucleon Fermi motion. Evaluating Eq. (47) using these smeared cross sections gives $\bar{d}(_{13}Al^{27}) = 0.142$, as compared with the value of 0.162 which results when nucleon motion is neglected. We see that the change in $\bar{d}$ is relatively small and is in the direction of reducing the size of charge-exchange effects; we expect these qualitative features to survive in a more careful treatment of nucleon-motion effects. In Table VIII we summarize the values of $\bar{d}(_{13}Al^{27})$ obtained in our original model and when various modifications are made.

C. Pion angular distributions

Up to this point we have only discussed charge-exchange corrections to cross sections in which the pion angular variables have been integrated out. Our model, however, makes specific predictions for angular distributions as well, and although they are much more subject to error than the integrated predictions,[21] they are essential for describing experimental situations in which the pion acceptance is limited. To describe the angular distribution predictions, we let the column vector

$$d\sigma(N_T\hat{q}) = \begin{pmatrix} d\sigma(N_T\hat{q}; +) \\ d\sigma(N_T\hat{q}; 0) \\ d\sigma(N_T\hat{q}; -) \end{pmatrix} \tag{49}$$

denote the free-nucleon-target pion-production cross section, with the pion emerging in direction $\hat{q}$. In the backward-forward scattering approximation, after undergoing nuclear interactions the

TABLE VI. Test of second averaging approximation for R' ($_{13}Al^{27}$).

$\sin^2\theta_W$	$\bar{r}(\nu_\mu \nu_\mu N_T)$	$\bar{r}(\nu_\mu \mu^- N_T)$	R' ($_{13}Al^{27}$) From Eq. (24)	From Eq. (15)
0	0.692	3.67	0.432	0.422
0.1	0.697	3.67	0.356	0.346
0.2	0.707	3.67	0.289	0.280
0.3	0.727	3.67	0.234	0.225
0.4	0.763	3.67	0.189	0.180
0.5	0.820	3.67	0.154	0.146
0.6	0.903	3.67	0.129	0.122
0.7	1.01	3.67	0.116	0.108
0.8	1.12	3.67	0.112	0.104
0.9	1.19	3.67	0.119	0.111
1.0	1.22	3.67	0.136	0.128

TABLE VII. Averaged charge-exchange parameters for various nuclei.

Nucleus	$\tilde{a}$	$\tilde{c}$	$\tilde{d}$
${}_5B^{10}$	0.846	0.0363	0.125
${}_6C^{12}$	0.811	0.0450	0.138
${}_7N^{14}$	0.790	0.0498	0.144
${}_8O^{16}$	0.807	0.0460	0.139
${}_{13}Al^{27}$	0.724	0.0642	0.162

pion can emerge either in direction $\hat{q}$ or with reversed direction $-\hat{q}$. In Appendix A, in addition to calculating the total expected fraction of emerging pions $f(\lambda, L)$, we also calculate the expected fractions $f_+(\lambda, L)$, $f_-(\lambda, L)$ which emerge, respectively, with or without a net change in direction. Using these to define a forward charge-exchange matrix M_+ and a backward matrix M_- in analogy with Eqs. (12), (37), and (39),

$$[M_\pm]=A_\pm\begin{pmatrix} 1-c_\pm-d_\pm & d_\pm & c_\pm \\ d_\pm & 1-2d_\pm & d_\pm \\ c_\pm & d_\pm & 1-c_\pm-d_\pm \end{pmatrix},$$

$$\begin{aligned} &A_\pm=g(W,k^2)a_\pm, \quad a_\pm=f_\pm(1)\\ &c_\pm=\tfrac{1}{3}-\tfrac{1}{2}f_\pm(\tfrac{5}{8})/f_\pm(1)+\tfrac{1}{6}f_\pm(\tfrac{1}{2})/f_\pm(1), \qquad (50)\\ &d_\pm=\tfrac{1}{3}[1-f_\pm(\tfrac{1}{2})/f_\pm(1)],\\ &f_\pm(\lambda)=\frac{\int_0^\infty bdb\, L(b)f_\pm[\lambda, L(b)]}{\int_0^\infty bdb\, L(b)}, \end{aligned}$$

we get for the charge-exchange-corrected pion angular distribution

$$d\sigma(T\hat{q})=[M_+]d\sigma(N_T\hat{q})+[M_-]d\sigma(N_T-\hat{q}). \qquad (51)$$

Since

$$[M_+]+[M_-]=[M], \qquad (52)$$

Eq. (51) implies that

$$d\sigma(T\hat{q})+d\sigma(T-\hat{q})=[M][d\sigma(N_T\hat{q})+d\sigma(N_T-\hat{q})], \qquad (53)$$

and so Eq. (51) reduces to our previous result for charge-exchange corrections when integrated over pion angle.

IV. CONCLUSIONS

We briefly summarize the results of the preceding sections, with particular emphasis on their implications for further experimental and theoretical work.

(1) Our model calculations confirm the suggestion of Perkins[2] that charge-exchange corrections to weak pion production are a substantial effect, even for relatively light nuclear targets. To improve our understanding of these corrections it is important to do the analogous pion electroproduction experiments on nuclear targets, both to implement the phenomenological procedures of Sec. II and to test the predictions of the detailed multiple-scattering model of Sec. III. In the context of the multiple-scattering model these electroproduction experiments have an independent nuclear physics interest, since they will permit a determination of the pion absorption cross section $\sigma_{abs}(W)$ entering into the Sternheim-Silbar[3] calculation, independent of assumptions about the magnitude of proton absorption in nuclear matter.

(2) Again, in the context of the multiple-scattering model, it is important to repeat the calculations of Sternheim and Silbar using the improved scattering approximation developed in Sec. II and Appendix A (as extended[9] to the case of a neutron excess). This will permit the extraction of an optimized pion absorption cross section $\sigma_{abs}(W)$ appropriate to the precise model which we use, and hopefully, may reduce some of the remaining areas of disagreement between the Sternheim-Silbar calculation and experiment.

(3) Our calculations suggest that the ratio $R'({}_{13}Al^{27})$ is larger than about 0.18 when the Weinberg parameter is in the currently interesting[2] range $\sin^2\theta_W \lesssim 0.35$. We do not attach great significance to the fact that this theoretical estimate of R' exceeds the upper bound of 0.14 reported by W. Lee,[1] since the discrepancy is easily of the order of uncertainties in the predictions of our production and charge-exchange models. We believe that a reasonably conservative statement is that if the hadronic neutral weak current has (up to isoscalar additions) the form of Eq. (43), and if $\sin^2\theta_W \lesssim 0.35$, then R' on an aluminum target is in the neighborhood of a 15% effect. Thus, an experiment capable of measuring R' to a level of a few percent will provide a decisive test of Eq. (43), and if Eq. (43) is correct, should permit a crude determination of $\sin^2\theta_W$.

Added note. A more recent calculation of proton-induced pion production on nuclear targets by

TABLE VIII. Effect of modifications of the model on $\tilde{d}({}_{13}Al^{27})$.

	No variations	Nucleon motion included	Pauli factors $h\to 1$	with $\frac{1}{2}\sigma_{abs}$	with $2\sigma_{abs}$
$\tilde{d}({}_{13}Al^{27})$	0.162	0.142	0.187	0.175	0.145

R. R. Silbar and M. M. Sternheim [Phys. Rev. C 8, 492 (1973)] gives a best-fit σ_{abs} given by

$$\sigma_{abs}(W)=\begin{cases} 30\ \mathrm{mb}\times\dfrac{T_\pi}{1.433M_\pi}, & T_\pi<1.433M_\pi \\ 51.3\ \mathrm{mb}\left(1-\dfrac{T_\pi}{3.455M_\pi}\right), & 1.433<T_\pi<3.455M_\pi\,. \end{cases} \tag{AN1}$$

Equation (AN1) is substantially larger than Eq. (27) in the region of low pion energy; Silbar and Sternheim attribute this difference largely to the inclusion of various nuclear corrections in their new calculation. Although it may not be consistent to use Eq. (AN1) in a model, such as ours, in which most of these nuclear corrections are neglected, we have nonetheless repeated the computation of Table II using Eq. (AN1), instead of Eq. (27), for σ_{abs}. The effect is to give values of R' which are about 17% larger than those tabulated in column 3 of Table II.

We have also repeated our calculations using the work of Ref. 9 to take the neutron excess in ${}_{13}\mathrm{Al}^{27}$ into account. The effect is to reduce R' by about 2.5% as compared with column 3 of Table II, indicating that the approximation of isotopic neutrality is a good one for ${}_{13}\mathrm{Al}^{27}$. This calculation also suggests that our neglect of changes in the nuclear isospin in the course of the multiple-scattering process (see Sec. III A 2) should cause an error of perhaps 10% at most in R'. Similarly, in analyzing the structure of M in Sec. II C we have implicitly neglected a possible change in the nuclear isospin arising from the pion production step (i.e., we continue to treat an initially $I=0$ nucleus as being in an $I=0$ state after the pion is produced); again, for nuclei which are not very light, the error resulting from making this approximation should be small and the three-parameter form given in Eq. (12) should give a reasonably good description of M.

ACKNOWLEDGMENTS

We wish to thank C. Baltay, M. A. B. Bég, K. M. Case, L. Hand, A. Kerman, B. W. Lee, W. Lee, H. J. Lipkin, S. B. Treiman, and J. D. Walecka for helpful conversations. One of us (S.L.A.) wishes to acknowledge the hospitality of the Aspen Center for Physics, where initial parts of this work were done.

APPENDIX A: ONE-DIMENSIONAL SCATTERING PROBLEM

In this appendix we solve the one-dimensional multiple-scattering problem on which our approximate solution for pion three-dimensional multiple scattering is based.[22] We briefly recapitulate the formulation of the problem given in the text. We consider a uniform one-dimensional medium extending from $x=0$ to $x=L$, in which pions are uniformly initially produced moving (say) to the right. The pions move in the medium with inverse interaction length κ, and at each interaction the pions are forward-scattered with probability μ_+ and back-scattered with probability μ_-, with a concomitant multiplication of the pion number by a factor of λ. The probabilities μ_+ and μ_- satisfy the constraint

$$\mu_+ + \mu_- \leq 1\,; \tag{A1}$$

when Eq. (A1) holds with the inequality, pion absorption is present. The problem is to find the expected numbers $f_\pm$ of pions eventually emerging from the medium either moving to the right (f_+: no over-all direction reversal) or to the left (f_-: over-all direction reversal), normalized to unit integrated initial pion density.

We begin by remarking that since f_+ (f_-) is even (odd) in the direction-reversal probability μ_-, it suffices to calculate

$$f=f_+ + f_-\,, \tag{A2}$$

the expected amplitude for pions to emerge in either direction. We then recover $f_\pm$ by splitting f into parts even and odd in μ_-. To formulate the multiple-scattering problem, we let $P(xj\,|\,yi)dx$ be the probability that a pion which after collision $n-1$ was at coordinate y moving in direction i ($i=l$, r = left, right) is, after collision n, in an interval dx at x moving in direction j. From the definitions of κ and $\mu_\pm$ given above, one easily finds that P, which does not depend on n, is given by

$$\begin{aligned} P(xr\,|\,yr) &= \mu_+\kappa e^{-\kappa(x-y)}\theta(x-y)\,,\\ P(xl\,|\,yr) &= \mu_-\kappa e^{-\kappa(x-y)}\theta(x-y)\,,\\ P(xl\,|\,yl) &= \mu_+\kappa e^{-\kappa(y-x)}\theta(y-x)\,,\\ P(xr\,|\,yl) &= \mu_-\kappa e^{-\kappa(y-x)}\theta(y-x)\,, \end{aligned} \tag{A3}$$

with θ the usual step function. Since the composition laws for *conditional probabilities* are the same as the quantum-mechanical composition laws for probability amplitudes, it is convenient to introduce a Dirac state notation by writing

$$\langle xj|P|yi\rangle = P(xj|yi);$$

$$\langle xj|P^2|yi\rangle = \int_0^L dz \sum_k \langle xj|P|zk\rangle\langle zk|P|yi\rangle, \quad \text{(A4)}$$

$$\langle xj|P^n|yi\rangle = \int_0^L dz \sum_k \langle xj|P|zk\rangle\langle zk|P^{n-1}|yi\rangle.$$

Letting $\rho^{(0)}(yi)$ be the initial density of produced pions moving in direction i, we then find that the density $\rho^{(n)}(xj)$ of pions which have undergone exactly n collisions and are moving in direction j is

$$\rho^{(n)}(xj) = \int_0^L dy \sum_i \langle xj|P^n|yi\rangle \rho^{(0)}(yi). \quad \text{(A5)}$$

The number of pions $N^{(n)}$ *emerging* from the medium after exactly n interactions is equal to the total number of pions present after n interactions less the number of such pions which interact once more in the medium,

$$N^{(n)} = \int_0^L dx[\rho^{(n)}(xl) + \rho^{(n)}(xr)] - \int_0^L dx\left[\int_0^x dz\,\kappa e^{-\kappa(x-z)}\rho^{(n)}(xl) + \int_x^L dz\,\kappa e^{-\kappa(z-x)}\rho^{(n)}(xr)\right]. \quad \text{(A6)}$$

Since each interaction multiplies the pion number by one factor of λ, the number $N^{(n)}$ must be weighted by λ^n in forming the expected number of pions leaving the medium. Taking $\rho^{(0)}(yi)$ to have the unit normalized value

$$\rho^{(0)}(yi) = \frac{1}{L}\delta_{i,r}, \quad \text{(A7)}$$

we get finally

$$f = \sum_{n=0}^{\infty} \lambda^n N^{(n)}. \quad \text{(A8)}$$

Equations (A3)–(A8) constitute the statement of our scattering problem. To write these equations more compactly, we introduce the additional notations

$$\langle zi|1|xj\rangle = \delta(z-x)\delta_{ij},$$
$$\langle z|P_{\rm tot}|xr\rangle = \kappa e^{-\kappa(z-x)}\theta(z-x), \quad \text{(A9)}$$
$$\langle z|P_{\rm tot}|xl\rangle = \kappa e^{-\kappa(x-z)}\theta(x-z),$$

in terms of which Eqs. (A5)–(A8) take the form

$$f = \int_0^L\int_0^L dz\,dx\left\{[\delta(z-x) - \langle z|P_{\rm tot}|xl\rangle]\sum_{n=0}^{\infty}\lambda^n\rho^{(n)}(xl) + [\delta(z-x) - \langle z|P_{\rm tot}|xr\rangle]\sum_{n=0}^{\infty}\lambda^n\rho^{(n)}(xr)\right\},$$

$$\sum_{n=0}^{\infty}\lambda^n\rho^{(n)}(xj) = \frac{1}{L}\int_0^L dy\left\langle xj\left|\sum_{n=0}^{\infty}\lambda^n P^n\right|yr\right\rangle = \frac{1}{L}\int_0^L dy\langle xj|(1-\lambda P)^{-1}|yr\rangle. \quad \text{(A10)}$$

Equation (A10) can be further simplified by noting that

$$\delta(z-x) - \langle z|P_{\rm tot}|xj\rangle = \left(1 - \frac{1}{\sigma_+ + \sigma_-}\right)\sum_i \langle zi|1|xj\rangle + \frac{1}{\sigma_+ + \sigma_-}\sum_i\langle zi|1-\lambda P|xj\rangle, \quad \text{(A11)}$$

with

$$\sigma_\pm = \lambda\mu_\pm. \quad \text{(A12)}$$

Substituting Eq. (A11) into Eq. (A10) we obtain, finally,

$$f = \left(1 - \frac{1}{\sigma_+ + \sigma_-}\right)\langle(1-\lambda P)^{-1}\rangle_{\rm av} + \frac{1}{\sigma_+ + \sigma_-}, \quad \text{(A13)}$$

with

$$\langle(1-\lambda P)^{-1}\rangle_{\rm av} = \frac{1}{L}\int_0^L\int_0^L dz\,dy\sum_i\langle zi|(1-\lambda P)^{-1}|yr\rangle. \quad \text{(A14)}$$

Equations (A12)–(A14) give a formal expression for f; to evaluate this expression explicitly we must determine the inverse operator appearing in Eq. (A14). Writing

$$\langle zi|(1-\lambda P)^{-1}|yj\rangle = \delta(z-y)\delta_{ij} + F(zi|yj) \quad \text{(A15)}$$

and defining

$$f(yj) = \int_0^L dz \sum_i F(zi|yj), \quad \text{(A16)}$$

we find that Eq. (A14) can be expressed in terms of $f(yj)$ as

$$\langle(1-\lambda P)^{-1}\rangle_{\rm av} = 1 + \frac{1}{L}\int_0^L dy\, f(yr)$$
$$= 1 + \frac{1}{L}\int_0^L dy\, f(yl), \quad \text{(A17)}$$

while the relation $(1-\lambda P)(1-\lambda P)^{-1} = 1$ implies that $f(yj)$ satisfies the integral equation

$$f(yj)=g(yj)+\int_0^L dz\sum_i f(zi)\langle zi|\lambda P|yj\rangle,$$
$$g(yj)=\int_0^L dz\sum_i\langle zi|\lambda P|yj\rangle. \tag{A18}$$

Referring back to Eq. (A3) for P, we easily see that $f(yj)$ and $g(yj)$ have the reflection symmetry

$$f(yl)=f(L-yr),\quad g(yl)=g(L-yr). \tag{A19}$$

Substituting Eq. (A3) into Eq. (A18) and using this symmetry, we find that Eq. (A18) reduces to the single integral equation

$$f(yl)=(\sigma_+ +\sigma_-)(1-e^{-\kappa y}) + \int_0^y dz[\kappa\sigma_+ f(zl)+\kappa\sigma_- f(L-zl)]e^{-\kappa(y-z)}. \tag{A20}$$

Multiplying Eq. (A20) by $e^{\kappa y}$ and differentiating, we find the equivalent differential equation and boundary condition

$$\kappa f(yl)+f'(yl)=(\sigma_+ +\sigma_-)\kappa+\kappa\sigma_+ f(yl) +\kappa\sigma_- f(L-yl),$$
$$f(0l)=0. \tag{A21}$$

The solution to Eq. (A21) has the form

$$f(yl)=\frac{\sigma_+ +\sigma_-}{1-(\sigma_+ +\sigma_-)}\left[1-\frac{h(y)}{h(0)}\right], \tag{A22}$$

with h a solution of the homogeneous equation

$$\kappa h(y)+h'(y)=\kappa\sigma_+ h(y)+\kappa\sigma_- h(L-y). \tag{A23}$$

To solve Eq. (A23), we try an exponential ansatz of the form

$$h(y)=e^{\kappa\sigma y}+\mu e^{-\kappa\sigma y}, \tag{A24}$$

which we find gives a solution when σ and μ are related to κ and $\sigma_\pm$ by

$$\sigma=[(1-\sigma_+)^2-\sigma_-^{\,2}]^{1/2},$$
$$\mu=\frac{\sigma+1-\sigma_+}{\sigma_-}e^{\kappa\sigma L}. \tag{A25}$$

It is now a matter of simple algebra to combine Eqs. (A13), (A17), (A22), (A24), and (A25) to give our final result for f, f_+, and f_-:

$$f=\frac{e^{\kappa\sigma L}-1}{\kappa\sigma L}\,\frac{1+\mu e^{-\kappa\sigma L}}{1+\mu}$$
$$=f_+ + f_-, \tag{A26}$$

$$f_+=\frac{e^{\kappa\sigma L}-1}{\kappa\sigma L}\,\frac{\mu^2 e^{-\kappa\sigma L}-1}{\mu^2-1},$$
$$f_-=\frac{e^{\kappa\sigma L}-1}{\kappa\sigma L}\,\frac{\mu(1-e^{-\kappa\sigma L})}{\mu^2-1}. \tag{A27}$$

As a check on Eq. (A27), we consider the special case in which there is no backward scattering, i.e., $\mu_-=0$. We find

$$f_-=0,$$
$$f_+=\frac{1-e^{-\kappa L(1-\lambda\mu_+)}}{\kappa L(1-\lambda\mu_+)}=\frac{1}{L}\int_0^L dy\, e^{-\kappa y(1-\lambda\mu_+)}, \tag{A28}$$

which is just the elementary exponential decay law appropriate to the case of forward propagation with effective absorption constant $\kappa(1-\lambda\mu_+)$, averaged over the length of the one-dimensional medium.

APPENDIX B: COMPARISON OF APPROXIMATE AND EXACT SCATTERING SOLUTIONS FOR A UNIFORM SPHERICAL GEOMETRY

In this appendix we calibrate the accuracy of the approximate scattering solution used in the text by comparing the approximate solution with the exact scattering solution in the case of a simple geometry. We consider a uniform sphere of radius R composed of material which scatters isotropically. Particles ("pions") are produced uniformly throughout the sphere and propagate with inverse interaction length κ. At each interaction the particles scatter isotropically, with the particle number simultaneously multiplied by a factor λ. We wish to find the expected number f of particles eventually emerging from the sphere, normalized to unit integrated initial particle density. We discuss successively the exact solution, two approximate solutions, and the numerical comparison.

1. Exact solution

The formulation of the solution to the spherical problem is closely analogous to the formulation of the one-dimensional problem in Appendix A, and we omit all details. Corresponding to Eqs. (A13), (A17), and (A18) we find[23]

$$f=\left(1-\frac{1}{\lambda}\right)\langle(1-\lambda P)^{-1}\rangle_{av}+\frac{1}{\lambda}, \tag{B1}$$

$$\langle(1-\lambda P)^{-1}\rangle_{av}=1+\frac{1}{\frac{4}{3}\pi R^3}\int_{|\vec{y}|\le R}d^3y f(\vec{y}), \tag{B2}$$

$$f(\vec{y})=g(\vec{y})+\int_{|\vec{z}|\le R}d^3z f(\vec{z})\lambda\frac{\kappa}{4\pi}\frac{e^{-\kappa|\vec{z}-\vec{y}|}}{|\vec{z}-\vec{y}|^2}, \tag{B3}$$

$$g(\vec{y})=\int_{|\vec{z}|\le R}d^3z\,\lambda\frac{\kappa}{4\pi}\frac{e^{-\kappa|\vec{z}-\vec{y}|}}{|\vec{z}-\vec{y}|^2}.$$

After spherical-averaging the scattering kernel, scaling out the sphere radius R, and expressing the solution of the integral equation in iterative form, we find

$$f = 1 + 3\left(1 - \frac{1}{\lambda}\right)\int_0^1 u^2 du \sum_{n=1}^{\infty} \lambda^n g^{(n)}(\rho, u),$$
$$g^{(0)} = 1,$$
$$g^{(n)}(\rho, u) = \rho\int_0^1 \tfrac{1}{2} v dv\, g^{(n-1)}(\rho, v)\,\frac{1}{u} \times [E_1(\rho|v-u|) - E_1(\rho(v+u))] \tag{B4}$$
$$E_1(x) = \int_x^{\infty} dt\,\frac{e^{-t}}{t},$$
$$\rho = \kappa R.$$

Since we are only interested in values of λ which are smaller than 1, the series in Eq. (B4) is convergent and f is readily calculated by repeated numerical integration.

2. Approximate solutions

We recall that the approximate scattering solution used in the text is obtained by projecting all forward- and backward-hemisphere scattering, respectively, onto the forward and backward directions, solving the resulting one-dimensional scattering problem as a function of optical thickness, and then integrating over the distribution of optical thickness actually present. For the spherical problem considered here substitution of Eqs. (A25) and (A26) into this recipe gives the following approximate formula for f:

$$f^{(1)} = \int_0^2 \tfrac{3}{8} u^2 du\,\frac{e^{\rho\sigma u} - 1}{\rho\sigma u}\,\frac{1 + \mu e^{-\rho\sigma u}}{1 + \mu};$$
$$\sigma = (1 - \lambda)^{1/2}, \tag{B5}$$
$$\mu = \frac{\sigma + 1 - \frac{1}{2}\lambda}{\frac{1}{2}\lambda}\,e^{\rho\sigma u},$$

which is readily evaluated by a single numerical integration. We also include in our comparison the scattering approximation used by Sternheim and Silbar, in which *all* scattering is projected onto the forward direction. In this case the relevant one-dimensional solution becomes the pure-forward-scattering solution of Eq. (A28) and we find a second approximate formula for f:

$$f^{(2)} = \int_0^2 \tfrac{3}{8} u^2 du\,\frac{e^{\rho(\lambda-1)u} - 1}{\rho(\lambda - 1)u}. \tag{B6}$$

3. Numerical comparison

Numerical results for f, $f^{(1)}$, and $f^{(2)}$ are given in Table IX for a wide range of values of λ and ρ. Agreement between the exact result f and the approximation $f^{(1)}$ used in the text is excellent over the entire range of parameters. The approximation $f^{(2)}$ used by Sternheim and Silbar is qualitatively correct, but develops significant deviations from the exact answer for large values of ρ. To interpret the parameter ρ in terms of nuclear size, we note that for a uniform spherical nucleus of radius $R \sim 1.3A^{1/3}$ F, and an interaction cross section characteristic of the peak of the (3, 3) resonance ($\sigma_{max} \sim 210$ mb $= 21$ F^2), we have

$$\rho \sim \frac{A}{\frac{4}{3}\pi R^3} \times \tfrac{2}{3}\sigma_{max} R$$
$$\sim 2A^{1/3}$$
$$\sim 6 \text{ for aluminum}$$
$$\sim 12 \text{ for lead.} \tag{B7}$$

Hence for aluminum our simple forward-backward approximation solves the multiple-scattering problem to an accuracy of better than 1%; even for the heaviest nuclei the approximation (with appropriate modifications to take neutron excess into account) should be good to better than 3%.

APPENDIX C: MISCELLANEOUS FORMULAS

We collect here the formulas for cross sections and Pauli factors used in the text.

1. Cross sections

For $\sigma_{\pi^+ p}(W)$ we use the simple form

$$\sigma_{\pi^+ p}(W) = \sigma_{(3,3)}(W) + 20 \text{ mb}, \tag{C1}$$

TABLE IX. Comparison of exact and approximate multiple-scattering solutions.

λ	ρ	f [Eq. (B4)]	$f^{(1)}$ [Eq. (B5)]	$f^{(2)}$ [Eq. (B6)]
0.5	0.5	0.827	0.827	0.835
	1	0.687	0.686	0.707
	2	0.489	0.488	0.527
	4	0.290	0.289	0.332
	8	0.154	0.153	0.182
	16	0.0790	0.0773	0.0930
0.6667	0.5	0.878	0.877	0.885
	1	0.766	0.764	0.789
	2	0.584	0.583	0.638
	4	0.370	0.368	0.445
	8	0.203	0.200	0.262
	16	0.105	0.102	0.138
0.8333	0.5	0.935	0.934	0.940
	1	0.867	0.865	0.885
	2	0.733	0.731	0.789
	4	0.524	0.522	0.638
	8	0.310	0.307	0.445
	16	0.166	0.161	0.262
0.9167	0.5	0.966	0.966	0.969
	1	0.928	0.927	0.940
	2	0.845	0.842	0.885
	4	0.678	0.676	0.789
	8	0.446	0.443	0.638
	16	0.251	0.244	0.445

with the first term the resonant cross section and the second term a constant approximation to the nonresonant background. (This formula slightly overestimates the cross section at and below the resonant peak, and underestimates it above resonance.) For $\sigma_{(3,3)}(W)$ we use the Roper parameterization,[24]

$$\sigma_{(3,3)}(W)=\sigma_{\max}\left(\frac{q_r}{q}\right)^2\frac{(\frac{1}{2}\Gamma)^2}{(q_0-q_{0r})^2+(\frac{1}{2}\Gamma)^2}, \qquad \text{(C2)}$$

with

$$\begin{aligned}
&q_0=\frac{W^2-M_N^{\,2}+M_\pi^{\,2}}{2W}, \quad q\equiv q(W)=(q_0^{\,2}-M_\pi^{\,2})^{1/2},\\
&q_{0r}=1.921M_\pi, \quad q_r=1.640M_\pi,\\
&\Gamma=\frac{1.262q^3/M_\pi}{(q_0+q_{0r})(1+0.504q^2/M_\pi^{\,2})},\\
&\sigma_{\max}=\frac{8\pi}{q_r^{\,2}}\approx 185\text{ mb}.
\end{aligned} \qquad \text{(C3)}$$

2. Pauli factors

We calculate the Pauli factors in the approximation of treating the nucleus as a collection of independent protons and neutrons with equal Fermi-sea radii $R_n=R_p=R\approx 1.6M_\pi$. Then the fraction of nucleons which can contribute, for given momentum transfer $\vec{\Delta}$ to the nucleus, is the fraction of the volume of a sphere of radius R centered at $\vec{0}$ which lies outside a second sphere of radius R centered at $\vec{\Delta}$. That is,[8]

$$h(W,\phi)=\begin{cases}\frac{3}{4}\eta-\frac{1}{16}\eta^3, & \eta\le 2\\ 1, & \eta\ge 2\end{cases} \qquad \text{(C4)}$$

with[25]

$$\eta=\frac{|\vec{\Delta}|}{R}\approx\frac{2q}{R}\sin(\tfrac{1}{2}\phi). \qquad \text{(C5)}$$

Performing the integrations over ϕ in Eq. (41), we find

$$\left.\begin{aligned}
h_+&=\alpha\frac{1}{\sqrt{2}}\frac{59}{70}-\alpha^3\frac{1}{\sqrt{2}}\frac{29}{420}\\
h_-&=\alpha\frac{136-59/\sqrt{2}}{70}-\alpha^3\frac{176-29/\sqrt{2}}{420}
\end{aligned}\right\}, \quad \alpha\le 1$$

$$\left.\begin{aligned}
h_+&=\alpha\frac{1}{\sqrt{2}}\frac{59}{70}-\alpha^3\frac{1}{\sqrt{2}}\frac{29}{420}\\
h_-&=2-\frac{4}{5}\alpha^{-2}+\frac{18}{35}\alpha^{-4}-\frac{4}{21}\alpha^{-6}-\alpha\frac{1}{\sqrt{2}}\frac{59}{70}+\alpha^3\frac{1}{\sqrt{2}}\frac{29}{420}
\end{aligned}\right\}, \quad 1\le\alpha\le\sqrt{2} \qquad \text{(C6)}$$

$$\left.\begin{aligned}
h_+&=1-\tfrac{4}{5}\alpha^{-2}+\tfrac{18}{35}\alpha^{-4}-\tfrac{4}{21}\alpha^{-6}\\
h_-&=1
\end{aligned}\right\}, \quad \sqrt{2}\le\alpha$$

with

$$\alpha=q/R. \qquad \text{(C7)}$$

For the production Pauli factor $g(W,k^2)$ we use the expression[26]

$$\begin{aligned}
&k_0=\frac{W^2-M_N^{\,2}-|k^2|}{2W}, \quad k=(k_0^{\,2}+|k^2|)^{1/2}\\
&g(W,k^2)=\frac{1}{2k}\left(\frac{3k^2+q^2}{2R}-\frac{5k^4+q^4+10k^2q^2}{40R^3}\right), \quad k+q\le 2R\\
&g(W,k^2)=\frac{1}{4qk}\left[(q+k)^2-\frac{4}{5}R^2-\frac{(k-q)^3}{2R}+\frac{(k-q)^5}{40R^3}\right], \quad k-q\le 2R\le k+q\\
&g(W,k^2)=1, \quad 2R\le k-q.
\end{aligned} \qquad \text{(C8)}$$

---

*Operated by Universities Research Association, Inc., under contract with the U. S. Atomic Energy Commission.

†Permanent address: Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel.

[1] B. W. Lee, Phys. Lett. 40B, 420 (1972); W. Lee, *ibid.* 40B, 423 (1972).

[2] D. H. Perkins, in *Proceedings of the XVI International Conference on High Energy Physics, Chicago-Batavia, Ill., 1972,* edited by J. D. Jackson and A. Roberts

(NAL, Batavia, Ill., 1973), Vol. 4, p. 189.

[3]M. M. Sternheim and R. R. Silbar, Phys. Rev. D 6, 3117 (1972). Earlier studies of pion charge exchange in nuclei involved the use of Monte Carlo techniques rather than analytical models. See N. Metropolis *et al.*, Phys. Rev. 110, 204 (1958); Yu. A. Batusov *et al.*, Yad. Fiz. 6, 158–164 (1967) [Sov. J. Nucl. Phys. 6, 116 (1968)]; C. Franzinetti and C. Manfredotti, CERN Report No. NPA/Int 67-30 (unpublished); C. Manfredotti, CERN Report No. NPA/Int 68-8 (unpublished).

[4]E. Fermi, Ric. Sci. 7 (2), 13 (1936); Report No. AECD-2664, 1951 (unpublished). For a discussion see E. Amaldi, in *Handbuch der Physik*, edited by S. Flügge (Springer, Berlin, 1959), Vol. 38, No. 2; G. M. Wing, Ref. 20.

[5]S. L. Adler, Ann. Phys. (N.Y.) 50, 189 (1968); Phys. Rev. D 9, 229 (1974).

[6]We adhere to the notations of Ref. 5 wherever possible.

[7]A detailed numerical calculation of the effect of Fermi motion on the production cross section indicates a substantial broadening of the resonance and a simultaneous shift of the resonance center to lower excitation energies. Both effects increase with increasing k^2. The effective *upper edge* of the resonance, however, is not shifted, and so an integration over experimental data from the effective threshold (which differs greatly from the threshold for pion production on nucleons at rest) to a fixed upper cutoff of

$$(k_0^L)_{\max} = \frac{(1.47\ \text{GeV})^2 - M_N^2 + |k^2|}{2M_N}$$

includes virtually the entire resonance. The area under the resonance obtained this way is essentially the same as the area obtained when Fermi motion is neglected. Hence, we expect production Fermi-motion effects to be relatively unimportant once the excitation energy has been integrated out, provided, of course, that one is not too close to a kinematic threshold for (3, 3)-resonance production.

[8]S. M. Berman, CERN report, 1961 (unpublished).

[9]This restriction is of course not necessary in principle. The extension of our multiple-scattering model to take a neutron excess into account will be given elsewhere [S. L. Adler, following paper, Phys. Rev. D 9, 2144 (1974)].

[10]Since Eq. (15) involves an integration over excitation energy k_0^L, we expect the Fermi-motion smearing of the production cross section to be relatively unimportant, and neglect it in the applications of Eq. (15) in Sec. III.

[11]The dominant resonant vector and axial-vector multipoles lead to different angular dependences of the production cross section.

[12]Some acceptance dependence in $\bar{A}$ could be tolerated, since it would tend to cancel between the numerator and denominator of R'.

[13]We are indebted to A. Kerman for a discussion about this point.

[14]When Pauli effects are included, the forward peak is washed out but the backward peak remains.

[15]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); *ibid.* 27, 1688 (1971); A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, edited by N. Svartholm (Almqvist, Stockholm, 1968), p. 367; G. 't Hooft, Nucl. Phys. B35, 167 (1971).

[16]This is strictly true only when $N_T = Z(p+n)$, whereas in the production calculation we have used $N_T = 13p + 14n$. The numerical effect of this change is small.

[17]C. H. Albright, B. W. Lee, E. A. Paschos, and L. Wolfenstein, Phys. Rev. D 7, 2220 (1973). Here G, θ_C, and α denote, respectively, the Fermi constant, the Cabibbo angle, and the fine-structure constant.

[18]The ratio 3.67 also includes the (small) effect of taking account of the actual n/p ratio in aluminum.

[19]The corresponding prediction for incident antineutrinos is 2.32.

[20]S. L. Adler, Ann. Phys. (N.Y.) 50, 189 (1968), Eq. (4E.7).

[21]Qualitatively, both nucleon Fermi motion and the deviations of the scattering angular distribution from pure "forward-backward scattering" would be expected to produce an angular smearing of the result of Eq. (51).

[22]For a nice pedagogical discussion of one-dimensional multiple scattering, see G. M. Wing, *An Introduction to Transport Theory* (Wiley, New York, 1962).

[23]The methods leading to these equationa are discussed in K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967). See especially Sec. 3.6.

[24]L. D. Roper, Phys. Rev. Lett. 12, 340 (1960).

[25]We have approximated $\bar{\Delta}$ by the isobar-frame momentum transfer. The approximation is bad only when η is so large that $h = 1$.

[26]Equation (C8) is obtained from Eq. (6C.6) of Ref. 18.

Erratum: Nuclear charge-exchange corrections to leptonic pion production in the (3,3)-resonance region [Phys. Rev. D 9, 2125 (1974)]

Stephen L. Adler, Shmuel Nussinov, and E. A. Paschos

Page 2138: In the second paragraph of the added note, the statement "The effect is to reduce R' by about 2.5%... should cause an error of perhaps 10% at most in R'" should be changed to read "The effect is to reduce R' by about 1%... should cause an error of at most a few percent in R'."

The following misprints should be corrected:

(i) Page 2127, Eq. (9b): The π_i to the right of the arrow should read π_f.

(ii) Page 2128, Eq. (15): The quantity $\sigma(\nu_\mu + T - \mu^- + T' + \pi^0)$ should read $\sigma(\nu_\mu + T - \mu^- + T'' + \pi^0)$.

(iii) Page 2134, Eq. (46): The quantity $\bar{r}(\nu_\mu \mu^- {}_{13}\mathrm{Al}^{27})$ should read $\bar{r}'(\nu_\mu \mu^- {}_{13}\mathrm{Al}^{27})$.

(iv) Page 2139, Eq. (A14): The quantity

$$\sum \text{ should read } \sum_i$$

(v) Page 2143: In Ref. 4 "G. M. Wing, Ref. 20" should read "G. M. Wing, Ref. 22"; in Ref. 26, "Eq. (6C.6) of Ref. 18" should read "Eq. (6C.6) of S. Adler, Ann. Phys. (N.Y.) 50, 189 (1966)."

VOLUME 33, NUMBER 25 PHYSICAL REVIEW LETTERS 16 DECEMBER 1974

Application of Current Algebra Techniques to Neutral-Current–Induced Threshold Pion Production

Stephen L. Adler
National Accelerator Laboratory, Batavia, Illinois 60510, and
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 7 October 1974)

I apply current-algebra techniques to study threshold pion production induced by the weak neutral current. In addition to specific predictions for the Weinberg-Salam–model current, I find upper bounds on the magnitude of threshold pion production for an isoscalar neutral current and for a general hadronic neutral current formed from the usual vector and axial-vector nonets. Violation of these bounds would suggest the presence of new coupling types in the neutral semileptonic interaction.

The initial experiments discovering weak neutral currents in high-energy deep-inelastic neutrino reactions[1] have now been supplemented with the observation of neutral-current effects in low-energy neutrino pion production.[2,3] Obtainable invariant-mass resolutions will permit the study of πN production in the threshold region below the (3,3) resonance, and in fact preliminary Argonne data[2] (without final corrections for neutron background) raise the possibility that the threshold cross section for $\pi^- p$ production by the neutral current may be appreciable. In this Letter we study threshold pion-production processes by using current-algebra, soft-pion techniques. I briefly describe the methods used in making such an analysis, and summarize the results obtained.

I begin by giving a simple analytic treatment of threshold pion production, which, although somewhat naive, illustrates the basic ideas which we exploit in our more careful numerical calculations. According to standard soft-pion lore,[4] the amplitude for the pion emission process $\mathcal{J}+\alpha \to \pi^j+\beta$, with α and β hadronic states and $\mathcal{J}$ an external current, is given as the sum of two terms. The first consists of a sum of external line insertions in which the pion π^j is emitted from the external hadronic lines of the pionless process $\mathcal{J}+\alpha\to\beta$, while the second is an equal-time commutator term proportional to the amplitude for the reaction $\mathcal{J}'+\alpha\to\beta$, with $\mathcal{J}'$ the modified current obtained from the commutator $\mathcal{J}'=[F_j^5,\mathcal{J}]$. In the case of neutral-current weak pion production, the current $\mathcal{J}$ is, of course, the hadronic weak neutral current and the states α and β are each a single free nucleon. For simplicity, let us restrict ourselves for the moment to cases in which the equal-time commutator term vanishes, as occurs, for example, if the current $\mathcal{J}$ is an isoscalar $V-A$ structure containing an arbitrary linear combination of $\mathcal{F}_0^{\lambda}$, $\mathcal{F}_8^{\lambda}$, $\mathcal{F}_0^{5\lambda}$, $\mathcal{F}_8^{5\lambda}$.[5] The pion emission amplitude then consists entirely of the external line insertion terms. Evaluating these terms at threshold (where the insertion on the outgoing nucleon line vanishes) and neglecting the pion mass in all kinematics, we find the following relation between threshold pion production and neutrino proton elastic scattering:

$$\frac{1}{|\vec{q}|}\frac{d\sigma(\nu+N\to\nu+N+\pi^j)}{d(k^2)dW}\bigg|_{\text{threshold}}=\frac{a^2}{4\pi^2M_\pi^2}\left(\frac{g_rM_\pi}{2M_N}\right)^2\frac{k^2}{M_N^2}\left(1+\frac{k^2}{4M_N^2}\right)\left(1+\frac{k^2}{2M_N^2}\right)^{-2}\frac{d\sigma(\nu+p\to\nu+p)}{d(k^2)}. \tag{1}$$

Here M_N, M_π are the nucleon and pion mass, W is the mass of the final π^jN isobar, $|\vec{q}|$ is the pion momentum in the isobaric rest frame, k^2 is the leptonic squared four-momentum transfer (spacelike, $k^2>0$), $g_r\approx 13.5$ is the pion-nucleon coupling constant, and the isospin matrix element a takes the values $|a|=\sqrt{2}$ for $\pi^j=\pi^\pm$ and $|a|=1$ for $\pi^j=\pi^0$. *The significance of Eq. (1) is that it allows one to translate an upper bound on the cross section for $\nu_\mu+p\to\nu_\mu+p$ into an upper bound on the strength of threshold pion production by the weak neutral current.*

As I have already suggested, the above derivation is too naive in a number of respects. First of all, the external line insertion terms are rapidly varying pole terms, and so the kinematic approximation of neglecting M_π in calculating them is dangerous. Secondly, by considering only cases in which the equal-time commutator term $\mathcal{J}'$ vanishes, we exclude from consideration such processes as π^- production in the SU(2) $\otimes$ U(1) gauge model. And finally, it is important to estimate the leading $O(q)$ corrections to the soft-pion approximation, and to calculate the effects in the threshold region of the tail of the

(3,3) resonance. We deal with these problems by using an extended version of a model for weak pion production which has been described in detail elsewhere.[6] In its original form, the model included the rapidly varying pole terms and the resonant (3,3) multipoles, with no kinematic approximations. The extensions consist of adding subtraction constants (in the dispersion-theory sense) to the non-Born terms of the model, which guarantee that it satisfies the relevant soft-pion theorems and which include the leading corrections (of first order in the pion four-momentum q and zeroth order in the lepton four-momentum transfer k) to the soft-pion limit. These latter corrections are calculated by the method of Low[7] and Adler and Dothan[7]; for the vector current amplitude they vanish, while for the isovector axial-vector amplitude they are related by partial conservation of axial-vector current to momentum derivatives of the pion-nucleon scattering amplitude at the crossing-symmetric point. For an isoscalar axial-vector current the order-q corrections cannot be precisely calculated, but a heuristic resonance-dominance argument suggests that they should be much smaller than in the isovector axial-vector case, and so we neglect them.

I give now the results of numerical calculations using the extended model in various cases, focusing attention on the reaction[8] $\nu_\mu + n \to \nu_\mu + \pi^- + p$.

(1) *Isoscalar neutral current.*—For the vector and axial-vector form factors in this case we take, for definiteness, a dipole formula with characteristic mass M_N,

$$F_1^S(k^2) = \lambda_1(1 + k^2/M_N^2)^{-2}, \quad 2M_N F_2^S(k^2) = \lambda_2(1 + k^2/M_N^2)^{-2}, \quad g_A^S(k^2) = \lambda_3(1 + k^2/M_N^2)^{-2}, \tag{2}$$

with λ_1, λ_2, and λ_3 free parameters. Assuming the 95% confidence bound[2]

$$\sigma(\nu_\mu + p \to \nu_\mu + p) \leq 0.32\sigma(\nu_\mu + n \to \mu^- + p), \tag{3}$$

we find that the cross section for $\nu_\mu + n \to \nu_\mu + \pi^- + p$, with $\pi^- p$ invariant mass W between[8] 1080 and 1120 MeV, is bounded by[9]

$$\sigma(\nu_\mu + n \to \nu_\mu + \pi^- + p) \leq 0.32\sigma(\nu_\mu + n \to \mu^- + p)[\sigma(\nu_\mu + n \to \nu_\mu + \pi^- + p)/\sigma(\nu_\mu + p \to \nu_\mu + p)], \tag{4a}$$

$$\leq 1.0 \times 10^{-41}\ \mathrm{cm}^2. \tag{4b}$$

The inequality in Eq. (4b) is obtained by maximizing the ratio in square brackets with respect to variation of λ_1, λ_2, and λ_3. We find in this case that the naive form of the low-energy theorem in Eq. (1) is reasonably good, predicting a bound about one-third as large as that of Eq. (4).[10]

(2) *Weinberg-Salam SU(2) ⊗ U(1) model.*—In the simplest, one-parameter version of this model, the neutral current has the form

$$\mathcal{J}_N^\lambda = \mathcal{F}_3^\lambda - \mathcal{F}_3^{5\lambda} - 2x(\mathcal{F}_3^\lambda + 3^{-1/2}\mathcal{F}_8^\lambda) + \Delta\mathcal{J}^\lambda, \quad x \equiv \sin^2\theta_W, \tag{5}$$

with $\Delta\mathcal{J}^\lambda$ an isoscalar, $V-A$, strangeness- and "charm"-current contribution which is conventionally assumed to couple only weakly to nonstrange low-mass hadrons. Neglecting $\Delta\mathcal{J}^\lambda$ for the moment, we can make an absolute calculation of the cross section for $\nu_\mu + n \to \nu_\mu + \pi^- + p$. We find, for $\pi^- p$ invariant mass W between 1080 and 1120 MeV, a predicted cross section of 0.75 $\times 10^{-41}$ cm^2. To assess the reliability of our calculations, Fig. 1 gives a comparison of our model with the Argonne National Laboratory results for the charged-current reaction $\nu_\mu + p \to \mu^- + \pi^+ + p$. The predicted cross section for $\pi^+ p$ invariant mass W between 1080 and 1120 MeV is 6.9 $\times 10^{-41}$ cm^2, in satisfactory agreement with the observed cross section of $(9.3 \pm 4.7) \times 10^{-41}$ cm^2.

In certain extensions of the original Weinberg-Salam model, the neutral current has the general form of Eq. (5), but with an adjustable strength parameter κ in front. A useful upper bound on the magnitude of κ is provided by deep-inelastic neutrino-scattering neutral-current data. In terms of the standard ratios $R_{\nu,\bar\nu} \equiv \sigma(\nu, \bar\nu + N \to \nu, \bar\nu + \Gamma)/\sigma(\nu, \bar\nu + N \to \mu^-, \mu^+ + \Gamma)$, we find[11] the 95% confidence limit[1]

$$1.5 \geq 3R_\nu + R_{\bar\nu} \geq \kappa^2[1 + (1 - 2x)^2]. \tag{6}$$

Continuing for the moment to neglect the isoscalar addition $\Delta\mathcal{J}^\lambda$, we can combine the bound of Eq. (6) with the extended model to predict that the cross section for neutral-current π^- production, with $\pi^- p$ invariant mass W between 1080 and 1120 MeV, is bounded by 1.5×10^{-41} cm^2 for all allowed values[12] of κ and x. Finally, we can include the isoscalar addition $\Delta\mathcal{J}^\lambda$ by parametriz-

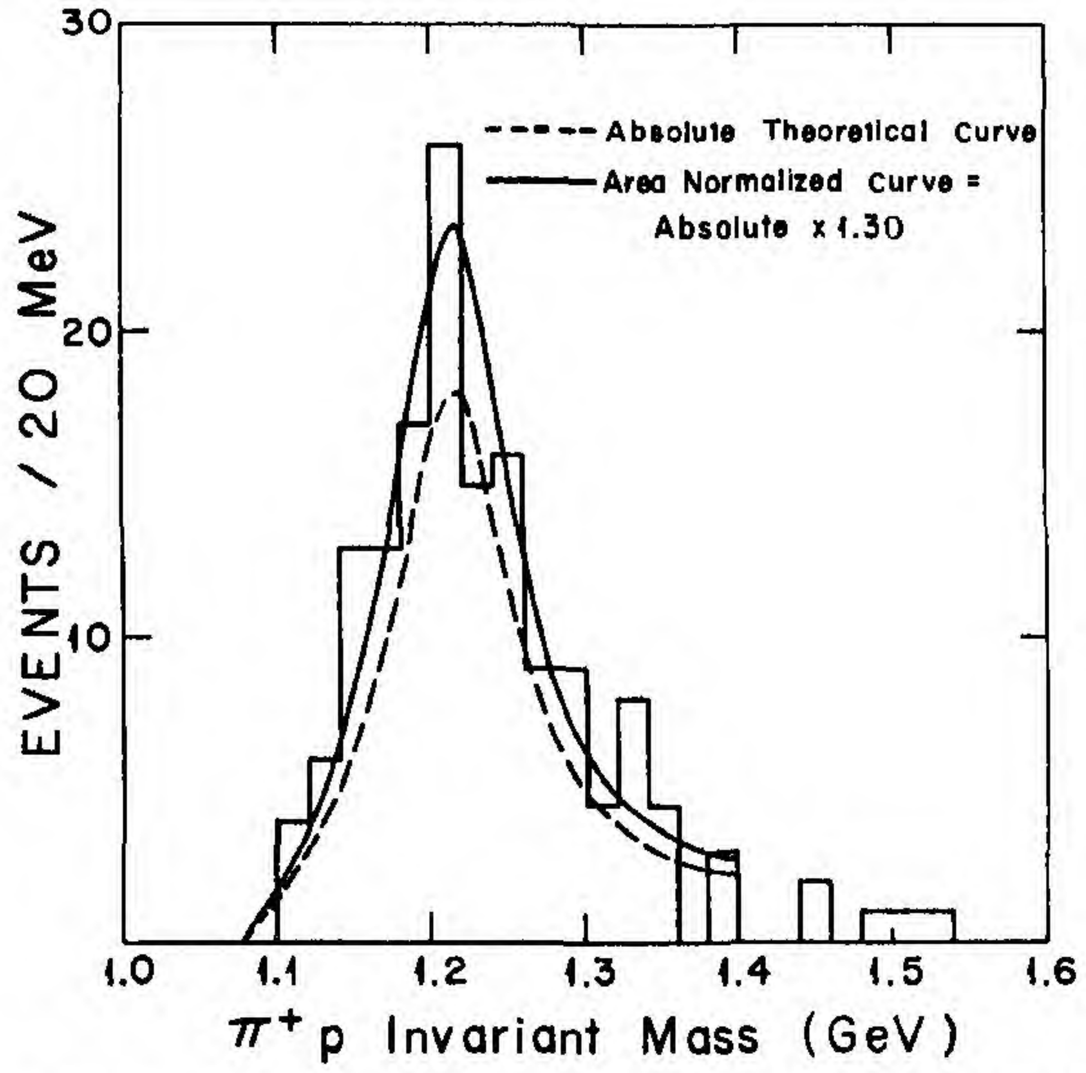


FIG. 1. Comparison of the extended pion production model with the Argonne National Laboratory charged-current data. Each event represents an Argonne flux-averaged cross section of 2.3×10^{-41} cm^2.

ing the total isoscalar contribution to $\mathcal{J}_N^{\lambda}$ as in Eq. (2), giving a cross section dependent on the five parameters κ, x, λ_1, λ_2, and λ_3. Combining the bounds of Eqs. (6) and (4a) with the extended model and maximizing over the five-parameter space,[12] we find that the cross section for $\nu_\mu + n \to \nu_\mu + \pi^- + p$, with W between 1080 and 1120 MeV, is bounded by 4.4×10^{-41} cm^2, for a general hadronic neutral current formed from the usual vector and axial-vector nonets.[13]

Experimental violation of this general bound, or the observation of evidence for an isoscalar neutral current together with violation of the bound of Eq. (4b), would suggest that the neutral current involves unusual types of coupling, in addition to or in place of the usually assumed $V-A$ structure. One possible source of violations could be an interaction of the $V-A$ type involving currents outside the usual quark-model vector and axial-vector nonets. An alternative source of violations could be the presence of S-, P-, and T-type neutral-current couplings.[14] If we define S, P, and T hadronic "currents" $\mathfrak{F}_j$, $\mathfrak{F}_j^{5}$, $\mathfrak{F}_j^{\lambda\sigma}$ and abstract their commutation relations from the quark-model forms

$$\mathfrak{F}_j = \bar{q}\tfrac{1}{2}\lambda_j q, \quad \mathfrak{F}_j^{5} = \bar{q}\tfrac{1}{2}\lambda_j\gamma_5 q,$$
$$\mathfrak{F}_j^{\lambda\eta} = \bar{q}\tfrac{1}{2}\lambda_j\sigma^{\lambda\eta}q, \tag{7}$$

then the commutator term $\mathcal{J}'$ appearing in the soft-pion analysis above will have SU(3) D- rather than F-type structure. This will substantially alter the structure of the low-energy theorems; for instance, the commutator term will no longer vanish for an isoscalar neutral current. The effect of this altered structure on the bounds given above is presently under study.

I wish to thank S. F. Tuan for stimulating discussions about the structure of neutral currents, S. B. Treiman for many helpful critical comments in the course of this work, and P. A. Schreiner and W. Y. Lee for conversations about the Argonne National Laboratory and Brookhaven National Laboratory neutrino experiments. I have also benefitted from discussions with R. F. Dashen, S. D. Drell, E. A. Paschos, and S. Weinberg.

---

[1]F. J. Hasert *et al.*, Phys Lett. 46B, 138 (1973); A. Benvenuti *et al.*, Phys. Rev. Lett. 32, 800 (1974).

[2]P. A. Schreiner, in Proceedings of the Seventeenth International Conference on High Energy Physics, London, England, July 1974 (to be published).

[3]Columbia-Rockefeller-Illinois Collaboration, in Proceedings of the Seventeenth International Conference on High Energy Physics, London, England, July 1974 (to be published).

[4]S. L. Adler and R. F. Dashen, *Current Algebras* (Benjamin, New York, 1968).

[5]Vanishing of the equal-time commutator in this case was noted by J. J. Sakurai, in Proceedings of the Fourth International Conference on Neutrino Physics and Astrophysics, Philadelphia, Pennsylvania, April 1974 (to be published).

[6]S. L. Adler, Ann. Phys. (New York) 50, 189 (1968). [See also S. L. Adler, Phys. Rev. D 9, 229 (1974).] The extended model is obtained by adding as subtraction constants Eq. (5A.21) for $\bar{A}_2^{(-)}|_0$, $\bar{A}_4^{(-)}|_0$, and $\bar{A}_7^{(+)}|_0$; Eq. (5A.22) for $\bar{V}_1^{(+)}|_0$, $\bar{V}_1^{(0)}|_0$, and $\bar{V}_6^{(-)}|_0$; Eq. (5A.9) for $\bar{A}_3^{(+)}|_0$; and Eq. (5A.30) for $\bar{A}_1^{(-)}|_0$. The order-q terms $\bar{A}_3^{(+)}|_0$ and $\bar{A}_1^{(-)}|_0$ were assumed to have k^2 dependence $(1+k^2/M_N^2)^{-2}$; variation of this assumed dependence produced only small changes in the results. We took the axial-vector form-factor mass as $M_A = 0.9$ GeV.

[7]F. E. Low, Phys. Rev. 110, 974 (1958); S. L. Adler and Y. Dothan, Phys. Rev. 151, 1267 (1966).

[8]Analogous bounds can be given for other pion-production channels and for larger invariant-mass intervals than the one considered here.

[9]The quoted bounds are not corrected for possible differences in the k^2 distributions of the reactions $\nu_\mu + p \to \nu_\mu + p$ and $\nu_\mu + n \to \mu^- + p$. For neutral-current form factors which decrease much more slowly than the charged-current form factors, the effect of such corrections would be to decrease the bounds.

[10]In the case of $\nu_\mu + N \to \nu_\mu + \pi^0 + N$ in the Weinberg-

Salam model, where Eq. (1) should formally hold, we find that the order-q corrections increase the (greatly suppressed) threshold pion production by an order of magnitude. As a result, the threshold π^0 production becomes comparable to that in $\nu_\mu + n \to \nu_\mu + \pi^- + p$ (where the order-q corrections have only an ~20% effect).

[11]Equation (6) assumes scaling, and also uses the fact that $\sigma(\bar{\nu}_\mu + N \to \mu^+ + \Gamma)/\sigma(\nu_\mu + N \to \mu^- + \Gamma) \approx \frac{1}{3}$. See A. Pais and S. B. Treiman, Phys. Rev. D 6, 2700 (1972).

[12]We search over all real values of x, even though only the range $0 \leq x \leq 1$ is physically meaningful in the SU(2)⊗U(1) model.

[13]This bound would be reduced if Eq. (6) were strengthened to include the isoscalar current contributions on the right-hand side.

[14]Tests for such couplings in the neutral current have been discussed by B. Kayser, G. T. Garvey, E. Fischbach, and S. P. Rosen (to be published) and by R. L. Kingsley, F. Wilczek, and A. Zee (to be published).

APPLICATION OF CURRENT ALGEBRA TECHNIQUES TO NEUTRAL-CURRENT-INDUCED THRESHOLD PION PRODUCTION. Stephen L. Adler [Phys. Rev. Lett. 33, 1511 (1974)].

An inadvertent confusion of meaning has resulted from replacement of commas by minus signs. On page 1511, column 2, and page 1513, column 1, "$V-A$" should read "V, A." Only on page 1512, column 1, was the "$V-A$" intended to mean "V minus A."

PHYSICAL REVIEW D VOLUME 12, NUMBER 9 1 NOVEMBER 1975

Application of current-algebra techniques to soft-pion production by the weak neutral current: V, A case *

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 21 April 1975)

We apply current-algebra techniques to study the constraints imposed on neutral-current-induced soft-pion production, using as input existing bounds on neutrino-proton elastic scattering and existing data on neutral-current-induced deep-inelastic scattering. In the case of a purely isoscalar weak neutral current, a simple soft-pion argument relates the cross section for threshold (in pion-nucleon invariant mass) weak pion production directly to the cross section for neutrino-proton elastic scattering. Hence, a bound on the latter cross section implies a bound on the former. To apply the method away from threshold and to nonisoscalar neutral currents, we extend a model which we had developed earlier for weak pion production in the (3,3) resonance region so as to include the low-energy-theorem constraints. Numerical work using the extended model shows that a threshold peak (now attributed to background) in preliminary Argonne data on $\nu+n\to\nu+p+\pi^-$ would have implied a threshold cross section much larger than can be obtained with any neutral current formed solely from members of the usual V, A nonets. We analyze recently reported Brookhaven National Laboratory results for neutral-current-induced soft-pion production under the simplifying assumption of a purely isoscalar V, A neutral current. We find in this case that the magnitude of the Brookhaven observations exceeds the theoretical maximum by more than a factor of 2 unless the assumed isoscalar current either contains a vector part with an anomalously large gyromagnetic ratio $|g|=|2M_NF_2/F_1|$ or involves the ninth (SU_3 singlet) axial-vector current. A vector part with a large $|g|$ value leads to characteristic modifications in the pion-nucleon invariant-mass spectrum for $M(\pi N)\leq 1.4$ GeV, an effect which should be testable in high-statistics experiments. Two other qualitative predictions of isoscalar V, A structures are (i) except for a narrow range of values of g, constructive V, A interference in $\nu+N\to\nu+N+\pi$ implies constructive interference in $\nu+p\to\nu+p$ and vice versa, and (ii) if V, A interference is observed in neutral weak processes then (as is well-known) the neutral interaction may make a parity-violating contribution to the pp, ep, and μp interactions. These features may help to distinguish V, A neutral-current couplings from alternative coupling types, which will be discussed in detail in subsequent papers of this series.

I. INTRODUCTION

The initial experiments discovering weak neutral currents in high-energy inclusive neutrino-nucleon scattering[1] have now been supplemented with the observation of neutral-current effects in the exclusive channel containing a pion-nucleon final state. Obtainable resolutions will permit the detailed study of pion-nucleon invariant-mass distributions in the region at and below the (3, 3) resonance. Some of the issues raised by recent Argonne National Laboratory (ANL)[2] and Brookhaven National Laboratory (BNL)[3] data on neutral-current exclusive channels are: (i) What is the expected magnitude for threshold (in invariant mass) neutral-current pion production? (ii) What are the implications if (3, 3) resonance excitation is not observed in neutral-current pion production? In the present paper we analyze these questions under the conventional assumption that the weak neutral current has a V, A spatial structure. A preliminary account of the analysis has appeared elsewhere.[4] In subsequent publications,[5] the same methods will be applied to the more general cases in which neutral-current couplings of S, P, T type appear, or in which V, A neutral currents with abnormal G parity are present.

The paper is organized as follows. In Sec. II we give a simple (although somewhat naive) analytic treatment of threshold pion production in the case when the neutral current is of pure isoscalar form, and use it to illustrate the methods employed in the more detailed treatments which follow. In Sec. III we develop the ingredients needed for a more elaborate treatment of bounds on soft-pion production. We first review the standard formulas describing neutrino-proton elastic scattering and deep-inelastic inclusive neutrino-nucleon scattering, the latter both in a general framework and within the context of quark-parton-model assumptions. We then describe the modifications needed to make our old dispersion-theoretic model for soft-pion production in the (3, 3) resonance region[6] consistent with all soft-pion theorem constraints, and discuss the inclusion of a well-defined set of corrections to the soft-pion limit. In Sec. IV we

give results of numerical studies of the pion production model, which show its validity in the charged current case. We then apply the formulas developed in Sec. III to a detailed numerical analysis of threshold pion production for the ANL neutrino flux case, considering a succession of more complex models for the structure of the neutral current, leading up to the most general neutral current which can be formed from members of the usual V,A nonets. We finally analyze low-invariant-mass $[W=M(\pi N)\leq 1.4\text{ GeV}]$ pion production for the BNL neutrino-flux spectrum, under the simplifying assumption of a pure isoscalar V,A neutral current. In Appendix A, we give the threshold low-energy theorem (analogous to that developed in Sec. II) which applies in the case of the SU(2)⊗U(1) model neutral current. In Appendix B, we attempt a rough estimate of the leading corrections to the soft-pion limit in the case of an isoscalar (octet) axial-vector current, and estimate the extent to which the corresponding corrections in the isovector current case are already included in the basic pion production model as a result of unitarization of the (3, 3) multipoles. In Appendix C, we discuss nuclear charge-exchange corrections for low-invariant-mass weak pion production and give a tabulation of the charge-exchange matrices for various nuclear targets of current theoretical interest.

II. SIMPLE ANALYTIC TREATMENT

We begin by giving a simple analytic treatment of threshold pion production, which, although somewhat naive, nonetheless illustrates the basic ideas exploited below in our more careful numerical calculations. The starting point for our derivation is the standard soft-pion formula[7] for pion emission in the process $\mathcal{J}+N\to\pi^j+N$, with $\mathcal{J}$ a general external current and N a nucleon. This reads[8]

$$\begin{aligned}\langle N(p_2)\,\pi(q)\,|\,\mathcal{J}(0)\,|\,N(p_1)\rangle = -\mathfrak{N}_{N\pi}\bar{u}(p_2)\Big[&\frac{g_r}{M_N g_A}J_j'(k-q)\\ &-\frac{g_r}{2M_N}\not{q}\gamma_5\tau_j\frac{\not{p}_2+M_N}{M_\pi^2+2p_2\cdot q}J(k)-J(k)\frac{\not{p}_1+M_N}{M_\pi^2-2p_1\cdot q}\frac{g_r}{2M_N}\not{q}\gamma_5\tau_j\\ &+\text{possible additional pion-pole "seagull" contribution}\Big]u(p_1)\,\psi_j^*+O(q)\,,\end{aligned}\tag{1}$$

with

$$\mathfrak{N}_N=\left(\frac{M_N}{p_{10}}\frac{M_N}{p_{20}}\right)^{1/2},\quad \mathfrak{N}_{N\pi}=\mathfrak{N}_N(2q_0)^{-1/2}\,,$$

$$\langle N(p_2)\,|\,\mathcal{J}(0)\,|\,N(p_1)\rangle\equiv\mathfrak{N}_N\bar{u}(p_2)J(p_2-p_1)u(p_1),\tag{2}$$

$$\langle N(p_2)\,|\,[F_j^5,\mathcal{J}(0)]\,|\,N(p_1)\rangle\equiv\mathfrak{N}_N\bar{u}(p_2)J_j'(p_2-p_1)u(p_1)\,.$$

In Eqs. (1) and (2), $k=p_2+q-p_1$ denotes the four-momentum carried by the external current, $g_r\approx 13.5$ is the pion-nucleon coupling constant, $g_A\approx 1.24$ is the nucleon axial-vector renormalization constant, $\bar{u}(p_2)$, $u(p_1)$ are nucleon spinors (including isospinors), and ψ_j is the isospin wave function of the emitted pion. The first term on the right-hand side of Eq. (1) is the usual equal-time commutator term which appears in soft-pion theorems, while the second and third terms are external-line-insertion terms in which the soft pion is emitted, respectively, from the final and initial nucleon lines. The additional pion-pole "seagull" piece[9] is necessary only when the pion-pole contributions of the first three terms do not add up to give the full pion-pole contribution expected for the reaction $\mathcal{J}+N\to\pi^j+N$.

Let us now specialize to the case of an isoscalar V,A external current $\mathcal{J}$, for which the equal-time commutator term vanishes[10] and for which there is no additional pion-pole "seagull" contribution. The entire soft-pion emission amplitude then comes from the external-line-insertion terms, which are most conveniently evaluated in the isobaric frame in which the final pion-nucleon system is at rest. In this frame, the insertion on the outgoing nucleon line vanishes at threshold in invariant mass, since when $\vec{p}_2=\vec{q}=0$ we have

$$\begin{aligned}\bar{u}(p_2)\not{q}\gamma_5(\not{p}_2+M_N)\,|_{\vec{p}_2=\vec{q}=0}&=q_0\bar{u}(p_2)\,\gamma_0\gamma_5(\not{p}_2+M_N)\,|_{\vec{p}_2=0}\\ &=0\,.\end{aligned}\tag{3}$$

So at threshold, for an isoscalar V,A current $\mathcal{J}$, the matrix element of Eq. (1) reduces to the single term

$$\begin{aligned}&\langle N(p_2)\,\pi(q)\,|\,\mathcal{J}(0)\,|\,N(p_1)\rangle\\ &\quad=\mathfrak{N}_{N\pi}\bar{u}(p_2)\,J(k)\frac{\not{p}_1+M_N}{M_\pi-2p_{10}}\frac{g_r}{2M_N}\gamma_0\gamma_5\tau_j u(p_1)\psi_j^*\,.\end{aligned}\tag{4}$$

On replacing the projection operator $(\not{p}_1+M_N)/2M_N$

by $\sum_s u(p_1 s)\,\bar{u}(p_1 s)$ and explicitly indicating the nucleon isospinors $\chi_2^\dagger, \chi_1$ and the helicity s_1 of the initial nucleon spinor $u(p_1)$, we find

$$\langle N(p_2)\,\pi(q)\,|\,\mathcal{J}(0)\,|\,N(p_1)\rangle = \sum \mathfrak{M}_{N\pi}\,\bar{u}(p_2)\,J(k)\,u(p_1 s)\,\frac{g_r}{M_\pi - 2p_{10}} \times[\bar{u}(p_1 s)\,\gamma_0\gamma_5\,u(p_1 s_1)]\,a\,, \quad (5)$$

with

$$a = \chi_2^\dagger\,\tau_j\,\chi_1\,\psi_j^* = \begin{cases} 1 \text{ for } p\to p+\pi^0 \\ -1 \text{ for } n\to n+\pi^0 \\ \sqrt{2} \text{ for } n\to p+\pi^- \\ \sqrt{2} \text{ for } p\to n+\pi^+ \end{cases} \quad (6)$$

The factor in square brackets in Eq. (5) is readily evaluated by using explicit expressions for the spinors, giving

$$\bar{u}(p_1 s)\,\gamma_0\gamma_5\,u(p_1 s_1) = \left(\frac{p_{10}+M_N}{2M_N}\right)\chi_s^\dagger\left(1,\ \frac{\vec{\sigma}\cdot\vec{k}}{p_{10}+M_N}\right) \times\begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix}\begin{pmatrix} 1 \\ -\vec{\sigma}\cdot\vec{k}/(p_{10}+M_N)\end{pmatrix}\chi_{s_1} = -\chi_s^\dagger\,\frac{\vec{\sigma}\cdot\vec{k}}{M_N}\,\chi_{s_1} = \frac{|\vec{k}|}{M_N}\,\delta_{s s_1}\,s_1\,, \quad (7)$$

where we have used the definition

$$-\vec{\sigma}\cdot\hat{k}\,\chi_{s_1} = s_1\chi_{s_1} \quad (8)$$

of the initial-state helicity. Since $s_1 = \pm 1$, this factor disappears when we square and sum over initial and final nucleon spins, so we get

$$\tfrac{1}{2}\sum_{N \text{ spins}} |\langle N(p_2)\,\pi(q)\,|\,\mathcal{J}(0)\,|\,N(p_1)\rangle|^2 \equiv \mathfrak{M}_{N\pi}^{\ 2}\,\langle|\mathfrak{M}_\pi|^2\rangle_{\text{threshold}} = \mathfrak{M}_{N\pi}^{\ 2}\left[\tfrac{1}{2}\sum_N\sum_{\text{spins}}|\bar{u}(p_2)\,J(k)\,u(p_1)|^2\right] \times\left(\frac{g_r}{M_\pi - 2p_{10}}\right)^2\frac{|\vec{k}|^2}{M_N^{\ 2}}\,a^2\,. \quad (9)$$

If we now make the approximation of neglecting the pion mass in all kinematics, the factor in square brackets in Eq. (9) becomes just the squared, spin-averaged matrix element $\langle|\mathfrak{M}|^2\rangle$ for νN elastic scattering, and Eq. (9) tells us that

$$\langle|\mathfrak{M}_\pi|^2\rangle_{\text{threshold}} = \langle|\mathfrak{M}|^2\rangle\left(\frac{g_r}{2M_N}\right)^2 a^2\,\frac{|\vec{k}|^2}{p_{10}^{\ 2}} = \langle|\mathfrak{M}|^2\rangle\left(\frac{g_r}{2M_N}\right)^2 a^2\,\frac{t}{M_N^{\ 2}}\,\frac{(1+t/4M_N^{\ 2})}{(1+t/2M_N^{\ 2})^2}\,,\quad t=-k^2\,. \quad (10)$$

Inserting phase-space factors according to

$$\frac{d\sigma(\nu+p\to\nu+p)}{dt} = \frac{1}{4\pi}\,\frac{1}{E^2}\,m_\nu^{\ 2}\langle|\mathfrak{M}|^2\rangle\,,$$
$$\frac{d\sigma(\nu+N\to\nu+N+\pi)}{dt\,dW} = \frac{1}{16\pi^3}\,\frac{|\vec{q}|}{E^2}\,m_\nu^{\ 2}\langle|\mathfrak{M}_\pi|^2\rangle\,, \quad (11)$$

with $|\vec{q}|$ the pion isobaric frame three-momentum, W the invariant mass of the final πN isobar, and E the initial lab neutrino energy, we get finally the relation

$$\frac{1}{|\vec{q}|}\,\frac{d\sigma(\nu+N\to\nu+N+\pi)}{dt\,dW}\bigg|_{\text{threshold}} = \frac{a^2}{4\pi^2 M_\pi^{\ 2}}\left(\frac{g_r M_\pi}{2M_N}\right)^2\frac{t}{M_N^{\ 2}}\left(1+\frac{t}{4M_N^{\ 2}}\right) \times\left(1+\frac{t}{2M_N^{\ 2}}\right)^{-2}\frac{d\sigma(\nu+p\to\nu+p)}{dt}\,, \quad (12)$$

$$a^2 = \begin{cases} 1 \text{ for } \pi^0 \text{ production} \\ 2 \text{ for } \pi^\pm \text{ production}\,. \end{cases}$$

We see that in the special case which has been under consideration, instead of obtaining a soft-pion relation between matrix elements, we obtain a relation directly in terms of reaction cross sections. *The significance of Eq. (12) is that it allows one to translate an upper bound on the strength of* $\nu+p\to\nu+p$ *into an upper bound on the strength of threshold pion production by the weak neutral current.*

As an illustration, let us apply Eq. (12) to the ANL data[2] by integrating over t and averaging over the ANL neutrino energy flux[11] $n_{\text{ANL}}(E)$, giving

$$\int dE\,n_{\text{ANL}}(E)\,\frac{M_N^{\ 2}}{|\vec{q}|}\,\frac{d\sigma(\nu+n\to\nu+p+\pi^-)}{dW} = \frac{2}{4\pi^2}\,\frac{M_N^2}{M_\pi^{\ 2}}\left(\frac{g_r M_\pi}{2M_N}\right)^2\int dE\,n_{\text{ANL}}(E)\int_0^{t_{\max}(E)} dt\,\frac{t}{M_N^{\ 2}}\left(1+\frac{t}{4M_N^{\ 2}}\right)\left(1+\frac{t}{2M_N^{\ 2}}\right)^{-2}\frac{d\sigma(\nu+p\to\nu+p)}{dt}\,. \quad (13)$$

We have multiplied both sides of Eq. (12) by M_N^2 so that they have the dimensions of a cross section; also for convenience, we assume the flux $n_{\mathrm{ANL}}(E)$ to be unit normalized,

$$\int dE\, n_{\mathrm{ANL}}(E) = 1\,, \tag{14}$$

so that we are considering flux-averaged cross sections. Using the ANL 95% confidence bound[12]

$$\sigma^{\mathrm{ANL}}(\nu+p\to\nu+p) \le 0.32\,\sigma^{\mathrm{ANL}}(\nu+n\to\mu^-+p) \approx 0.25\times10^{-38}\ \mathrm{cm}^2\,, \tag{15}$$

and assuming the t dependence of the charged-current quasielastic and neutral-current elastic cross sections to be similar,[13,14] we find that the right-hand side of Eq. (13) is bounded by $0.32\times0.46\times10^{-38}\ \mathrm{cm}^2 \approx 0.15\times10^{-38}\ \mathrm{cm}^2$. Using 20-MeV invariant-mass bins, we can then estimate a bound on the flux-averaged cross section in the two bins nearest threshold as shown in Table I, giving the result

$$\sigma^{\mathrm{ANL}}_{2\ \mathrm{bin}} \equiv \sigma(\nu+n\to\nu+p+\pi^-\ \text{ANL flux averaged, } 1.08\ \mathrm{GeV}\le W\le 1.12\ \mathrm{GeV}) \le 0.6\times10^{-41}\ \mathrm{cm}^2\,. \tag{16}$$

For comparison, the preliminary ANL data on $\nu+n\to\nu+p+\pi^-$, before final background subtraction, showed ~5 events in the first two bins, which would have corresponded to a cross section of

$$\sigma^{\mathrm{ANL}}_{2\ \mathrm{bin}}\ \text{(before background subtraction)} \sim 20\times10^{-41}\ \mathrm{cm}^2\,, \tag{17}$$

in strong violation of the bound of Eq. (16). It is now considered very probable that these events do *not* represent a true neutral-current effect, but arise from various neutron-induced backgrounds.

As we have already remarked, the above treatment is too naive in a number of respects. First of all, the restriction to cases, such as[10] that of an isoscalar neutral current $\mathcal{S}$, for which the equal-time commutator term vanishes excludes from consideration such processes as π^- production in the SU(2) ⊗ U(1) gauge model. Secondly, the external-line-insertion terms are rapidly varying pole terms, and so the kinematic approximation of neglecting M_π in calculating them can be dangerous. Finally, it is important to estimate the leading $O(q)$ corrections to the soft-pion approximation, and to calculate the effects in the threshold region of the tail of the (3, 3) resonance. As discussed in detail in Sec. III, we deal with these problems by using an extended version of a model for the weak pion-production amplitude which we have described elsewhere.[6] The extension will permit us to study the entire low-invariant-mass region $W \le 1.4$ GeV, rather than just the first 40 MeV or so around threshold. For completeness, however, we give in Appendix A the analog of the threshold low-energy theorem of Eq. (12) for the case of the SU(2) ⊗ U(1)-model neutral current. The formulas of Appendix A still neglect the pion mass in the kinematics [as well as the leading $O(q)$ corrections and the (3, 3) resonance tail] and are not used in the subsequent numerical work.

III. DETAILED TREATMENT

We proceed in this section to set out the basis for a more detailed numerical treatment of bounds on weak pion production by a V, A weak neutral current. The basic idea, as developed above, is to use soft-pion techniques to relate weak pion production to elastic neutrino-proton scattering, and to use experimental bounds on the latter. It will also be useful, at some stages of the analysis, to impose constraints obtained from experimental data[1,15] on deep-inelastic inclusive neutrino-nucleon scattering induced by the weak neutral current. In Sec. III A we give the necessary vertex structure and cross-section formulas needed to describe neutrino-nucleon elastic scattering. In Sec. III B we give the necessary formulas for using deep-inelastic information; first, in a rather general form assuming only scaling and the fact that

$$\sigma(\bar\nu+N\to\mu^++\Gamma)/\sigma(\nu+N\to\mu^-+\Gamma)\approx\tfrac{1}{3}\,,$$

TABLE I. Application of Eq. (13) to bound the cross section for $\nu+n\to\nu+p+\pi^-$ in the two ANL 20-MeV bins nearest invariant-mass threshold.

W at bin center	$\lvert\vec{q}\rvert/M_N$	dW/M_N	Bound on right-hand side of Eq. (13)	Bound on cross section in bin $[(\lvert\vec{q}\rvert dW/M_N^2)\times0.15\times10^{-33}\ \mathrm{cm}^2]$
1.09	6.4×10^{-2}	0.021	$0.15\times10^{-38}\ \mathrm{cm}^2$	$0.20\times10^{-41}\ \mathrm{cm}^2$
1.11	1.1×10^{-1}	0.021	$0.15\times10^{-33}\ \mathrm{cm}^2$	$0.35\times10^{-41}\ \mathrm{cm}^2$

and then in a more restrictive form which makes use of quark-parton model and quark-model assumptions. Finally, in Sec. III C we develop the extended weak-pion-production model which remedies the defects in our naive treatment enumerated at the end of Sec. II.

A. Elastic neutrino-nucleon scattering

We shall consider in what follows the most general V,A weak neutral current which can be formed from members of the usual vector and axial-vector nonets. We write for the neutral-current effective Lagrangian

$$\mathcal{L}_{\rm eff}^{N}=\frac{G}{\sqrt{2}}\,\bar{\nu}\gamma_\lambda(1-\gamma_5)\,\nu\mathcal{J}_N^\lambda,$$
$$\mathcal{J}_N^\lambda=g_{V0}\,\mathcal{F}_0^\lambda+g_{V3}\,\mathcal{F}_3^\lambda+g_{V8}\,\mathcal{F}_8^\lambda-g_{A0}\,\mathcal{F}_0^{5\,\lambda}-g_{A3}\,\mathcal{F}_3^{5\,\lambda}-g_{A8}\,\mathcal{F}_8^{5\,\lambda}, \tag{18}$$

with $\mathcal{F}_j^\lambda$, $\mathcal{F}_j^{5\lambda}$ nonet currents represented in the quark model (with quark field ψ) by[16]

$$\mathcal{F}_j^\lambda=\bar{\psi}\gamma^\lambda\tfrac{1}{2}\lambda_j\psi\,,$$
$$\mathcal{F}_j^{5\,\lambda}=\bar{\psi}\gamma^\lambda\gamma_5\tfrac{1}{2}\lambda_j\psi\,. \tag{19}$$

We express the nucleon matrix elements of the neutral members of these current nonets in the form[16]

$$\langle N(p_2)|\,\mathcal{F}_j^\lambda\,|N(p_1)\rangle=\mathfrak{N}_N\bar{u}(p_2)[F_1^{(j)}(k^2)\gamma^\lambda+iF_2^{(j)}(k^2)\sigma^{\lambda\eta}k_\eta]\,t_j u(p_1)\,,$$
$$\langle N(p_2)|\,\mathcal{F}_j^{5\,\lambda}\,|N(p_1)\rangle=\mathfrak{N}_N\bar{u}(p_2)[g_A^{(j)}(k^2)\gamma^\lambda\gamma_5+h_A^{(j)}(k^2)k^\lambda\gamma_5]\,t_j u(p_1)\,, \tag{20}$$
$$t_0=\tfrac{1}{2}(\tfrac{2}{3})^{1/2},\quad t_3=\tfrac{1}{2}\tau_3,\quad t_8=\tfrac{1}{2}(\tfrac{1}{3})^{1/2}\,.$$

The vector and axial-vector form factors defined in Eq. (20) are related to the standard nucleon form factors $F_{1,2}^{V,S}(k^2)$, $g_A(k^2)$, $h_A(k^2)$ by

$$F_{1,2}^{(3)}(k^2)=F_{1,2}^{V}(k^2),\quad g_A^{(3)}(k^2)=g_A(k^2)\,,$$
$$F_{1,2}^{(8)}(k^2)=3F_{1,2}^{S}(k^2),\quad h_A^{(3)}(k^2)=h_A(k^2)\,. \tag{21}$$

Defining total form factors $F_{1,2}^T(k^2)$, $g_A^T(k^2)$ by

$$F_{1,2}^T(k^2)=\tfrac{1}{2}(\tfrac{2}{3})^{1/2}g_{V0}F_{1,2}^{(0)}(k^2)+\tfrac{1}{2}\epsilon g_{V3}F_{1,2}^{(3)}(k^2)+\tfrac{1}{2}(\tfrac{1}{3})^{1/2}g_{V8}F_{1,2}^{(8)}(k^2)\,,$$
$$g_A^T(k^2)=\tfrac{1}{2}(\tfrac{2}{3})^{1/2}g_{A0}g_A^{(0)}(k^2)+\tfrac{1}{2}\epsilon g_{A3}g_A^{(3)}(k^2)+\tfrac{1}{2}(\tfrac{1}{3})^{1/2}g_{A8}g_A^{(8)}(k^2)\,, \tag{22}$$

with $\epsilon=1$ for $\nu+p\to\nu+p$ and $\epsilon=-1$ for $\nu+n\to\nu+n$, the differential cross section for neutrino-nucleon scattering takes the form

$$\frac{d\sigma(\nu+N\to\nu+N)}{dt}=\frac{G^2}{8\pi E^2M_N^2}\{[\,|F_1^T|^2+|g_A^T|^2+t|F_2^T|^2][4M_N^2E^2-t(M_N^2+2M_NE)]$$
$$+\tfrac{1}{2}t[\,|g_A^T|^2(t+4M_N^2)+|F_1^T+2M_NF_2^T|^2t]+\mathrm{Re}[g_A^{T*}(F_1^T+2M_NF_2^T)]\,t(4M_NE-t)\}\,. \tag{23}$$

For incident antineutrinos, the sign of g_A^T in Eq. (23) is reversed.

B. Deep-inelastic inclusive neutrino-nucleon scattering

We turn next to the constraints on the coefficients appearing in Eq. (18) which are imposed by experimental measurements[1,15] of the deep-inelastic inclusive neutrino-nucleon scattering ratios[8]

$$R_\nu\equiv\sigma(\nu+N\to\nu+\Gamma)/\sigma(\nu+N\to\mu^-+\Gamma)\,,$$
$$R_{\bar\nu}\equiv\sigma(\bar\nu+N\to\bar\nu+\Gamma)/\sigma(\bar\nu+N\to\mu^++\Gamma)\,. \tag{24}$$

The charged-current-induced reactions in the denominators in Eq. (24) are described by the usual charged-current effective Lagrangian

$$\mathcal{L}_{\rm eff}^{\rm ch}=\frac{G}{\sqrt{2}}\,\bar{\mu}\gamma_\lambda(1-\gamma_5)\,\nu\mathcal{J}_{\rm ch}^\lambda+\text{adjoint}\,, \tag{25}$$

with

$$\mathcal{J}_{\rm ch}^\lambda=\cos\theta_C(\mathcal{F}_{1+i2}^\lambda-\mathcal{F}_{1+i2}^{5\,\lambda})+\sin\theta_C(\mathcal{F}_{4+i5}^\lambda-\mathcal{F}_{4+i5}^{5\,\lambda})\,. \tag{26a}$$

In what follows we aim only at getting formulas which hold to an accuracy of 10 or 20%, and so we make at the outset the approximation of taking the Cabibbo angle θ_C to be zero, which simplifies Eq. (26a) to read

$$\mathcal{J}_{\rm ch}^\lambda\approx\mathcal{F}_{1+i2}^\lambda-\mathcal{F}_{1+i2}^{5\,\lambda}\,. \tag{26b}$$

The virtue of using Eq. (26b) is that the vector and axial-vector parts of the charged current are then related by an isospin rotation to the corre-

sponding isovector vector and axial-vector terms in the neutral current of Eq. (18).

To proceed with the analysis, we assume the validity of Bjorken scaling[17] in deep-inelastic charged-current and neutral-current-induced inclusive neutrino reactions. Considering for the moment the charged-current cross sections $\sigma(\nu+N\to\mu^-+\Gamma)$ and $\sigma(\overline{\nu}+N\to\mu^++\Gamma)$, we review a standard analysis[18] starting from the formula

$$\frac{\sigma(\overline{\nu}+N\to\mu^++\Gamma)}{\sigma(\nu+N\to\mu^-+\Gamma)}=\frac{\int_0^1 dx\,a_S+\frac{1}{3}\int_0^1 dx\,x\,a_L+\int_0^1 dx\,x\,a_R}{\int_0^1 dx\,a_S+\int_0^1 dx\,x\,a_L+\frac{1}{3}\int_0^1 dx\,x\,a_R}\,, \tag{27}$$

where

$$a_S=\tfrac{1}{2}F_2-xF_1\geq 0\,,$$
$$a_L=F_1-\tfrac{1}{2}F_3\geq 0\,,$$
$$a_R=F_1+\tfrac{1}{2}F_3\geq 0\,,$$
$$x=1/\omega=\text{scaling variable}\,,$$

with $F_{1,2,3}$ the deep-inelastic structure functions in the scaling limit, and where an average nucleon target $N=\frac{1}{2}(n+p)$ has been assumed. Empirically, it is found that

$$\sigma(\overline{\nu}+N\to\mu^++\Gamma)/\sigma(\nu+N\to\mu^-+\Gamma)\approx\tfrac{1}{3}\,,$$

which implies that $a_S\approx a_R\approx 0$, that is,

$$F_3(x)\approx-2F_1(x),\quad F_2(x)\approx 2xF_1(x)\,. \tag{28}$$

Splitting F_1 and F_2 into vector and axial-vector contributions

$$F_{1,2}(x)=F^V_{1,2}(x)+F^A_{1,2}(x)\,, \tag{29}$$

we may rewrite the relations of Eq. (28) as

$$\begin{aligned}&\tfrac{1}{4}|F_3(x)|\approx\tfrac{1}{2}[F^V_1(x)+F^A_1(x)]\,,\\ &F^V_2(x)+F^A_2(x)\approx 2x[F^V_1(x)+F^A_1(x)].\end{aligned} \tag{30}$$

Comparing Eq. (30) with the Schwarz inequality[19]

$$\tfrac{1}{4}|F_3(x)|\leq[F^V_1(x)F^A_1(x)]^{1/2}\leq\tfrac{1}{2}[F^V_1(x)+F^A_1(x)] \tag{31}$$

and the positivity inequalities

$$\begin{aligned}&F^V_2(x)\geq 2xF^V_1(x)\,,\\ &F^A_2(x)\geq 2xF^A_1(x)\,,\end{aligned} \tag{32}$$

we learn that

$$\begin{aligned}F^V_1(x)&\approx F^A_1(x)\approx\frac{1}{2x}F^V_2(x)\approx\frac{1}{2x}F^A_2(x)\\ &\approx-\tfrac{1}{4}F_3(x)\approx\tfrac{1}{2}F_1(x)\approx\frac{1}{4x}F_2(x)\,.\end{aligned} \tag{33}$$

Now let us turn our attention to the deep-inelastic ratios of Eq. (24). If we again take N to be an average nucleon target, the isovector and isoscalar terms in the neutral current of Eq. (18) do not interfere, and so we get a lower bound on R_ν and $R_{\overline{\nu}}$ by neglecting the isoscalar contributions to the cross section. Using the fact, already mentioned, that the isovector pieces of Eq. (18) are related to the corresponding isovector pieces of Eq. (26) by an isospin rotation, we find that

$$\begin{pmatrix}R_\nu\\ R_{\overline{\nu}}\end{pmatrix}\geq\frac{1}{2}\,\frac{\int_0^1 dx\{g_{V3}^2[\frac{1}{3}xF^V_1(x)+\frac{1}{2}F^V_2(x)]+g_{A3}^2[\frac{1}{3}xF^A_1(x)+\frac{1}{2}F^A_2(x)]\mp g_{V3}g_{A3}\frac{1}{3}xF_3(x)\}}{\int_0^1 dx[\frac{1}{3}xF_1(x)+\frac{1}{2}F_2(x)\mp\frac{1}{3}xF_3(x)]}\,. \tag{34}$$

Substituting now the relations of Eq. (33), we get the simple inequalities[20]

$$\begin{aligned}&R_\nu\geq\tfrac{1}{6}(g_{V3}^2+g_{A3}^2+g_{V3}g_{A3})\,,\\ &R_{\overline{\nu}}\geq\tfrac{1}{2}(g_{V3}^2+g_{A3}^2-g_{V3}g_{A3})\,.\end{aligned} \tag{35}$$

When added in the linear combination $3R_\nu+R_{\overline{\nu}}$ which eliminates the vector–axial-vector interference term, and combined with 95% confidence limits inferred from current measurements of R_ν and $R_{\overline{\nu}}$, the inequalities of Eq. (35) yield the constraint

$$1.5\geq 3R_\nu+R_{\overline{\nu}}\geq g_{V3}^2+g_{A3}^2\,, \tag{36}$$

which will be used in our subsequent analysis. As we have already emphasized, in getting Eq. (36) we have only used the assumption of scaling together with the empirical observation of a charged-current antineutrino-to-neutrino inclusive cross-section ratio of $\approx\frac{1}{3}$.

In order to strengthen Eq. (36) so as to include the isoscalar current terms in Eq. (18), it is necessary to go beyond the assumptions just stated by using information from the quark-parton model. Specifically, we will make use of the standard spin-$\frac{1}{2}$ quark-parton model for deep-inelastic scattering,[21] with the additional assumptions that the strange parton and the antiparton content of the nucleon may be neglected[22] [the latter of these assumptions is suggested by the approximate relations of Eq. (33)]. The quark-parton model in this form is expected[22] to be good to an accuracy of order 20%, and has the great virtue that all x dependence (for an average nucleon target) appears in a single universal over-all factor which drops out in the ratios $R_{\nu,\overline{\nu}}$. A straightforward calculation then gives

$$\begin{pmatrix} R_\nu \\ \frac{1}{3}R_{\bar\nu} \end{pmatrix} = \tfrac{2}{3}\left\{\left[g_{V0}\tfrac{1}{2}(\tfrac{2}{3})^{1/2} + g_{V6}\tfrac{1}{2}(\tfrac{1}{3})^{1/2}\right]^2 + (\tfrac{1}{2}g_{V3})^2 + \left[g_{A0}\tfrac{1}{2}(\tfrac{2}{3})^{1/2} + g_{A6}\tfrac{1}{2}(\tfrac{1}{3})^{1/2}\right]^2 + (\tfrac{1}{2}g_{A3})^2\right\}$$
$$\pm \tfrac{2}{3}\left\{\left[g_{V0}\tfrac{1}{2}(\tfrac{2}{3})^{1/2} + g_{V8}\tfrac{1}{2}(\tfrac{1}{3})^{1/2}\right]\left[g_{A0}\tfrac{1}{2}(\tfrac{2}{3})^{1/2} + g_{A8}\tfrac{1}{2}(\tfrac{1}{3})^{1/2}\right] + \tfrac{1}{2}g_{V3}\tfrac{1}{2}g_{A3}\right\}, \tag{37}$$
$$1.5 \geq R_{\bar\nu} + 3R_\nu = \left[g_{V0}(\tfrac{2}{3})^{1/2} + g_{V8}(\tfrac{1}{3})^{1/2}\right]^2 + g_{V3}{}^2 + \left[g_{A0}(\tfrac{2}{3})^{1/2} + g_{A8}(\tfrac{1}{3})^{1/2}\right]^2 + g_{A3}{}^2 .$$

One additional piece of information which will be needed, in order to use the constraint of Eq. (37) in an analysis of low-energy pion production, is knowledge of the renormalization constants $g_A^{(0,8)}(0)$ and $F_2^{(0,8)}(0)$ which describe the one-nucleon matrix elements of the isoscalar currents appearing in Eq. (18). The constant $g_A^{(8)}(0)$ is fairly reliably fixed by SU(3) to have the value[23]

$$g_A^{(8)}(0) \approx (3 - 4\times 0.66)\,1.24 = 0.45 , \tag{38a}$$

while the measured value of $F_2^{(8)}(0)$ is

$$2M_N F_2^{(8)}(0)/F_1^{(8)}(0) = -0.12 . \tag{38b}$$

For the constants $g_A^{(0)}(0)$ and $F_2^{(0)}(0)$ recourse must be made to a quark-model analysis of current-renormalization constants,[16] which gives[24]

$$g_A^{(0)}(0) \approx \tfrac{3}{5}\,1.24 = 0.74 ,$$
$$2M_N F_2^{(0)}(0)/F_1^{(0)}(0) \approx -0.1 \tag{39}$$

for the unitary-singlet renormalization constants.

C. Extended model for weak pion production

We turn finally to a description of the extended model for weak pion production which we will use in the numerical calculations of Sec. IV. As an aid to the discussion which follows, let us first rewrite the pion-production matrix element of Eq. (1) in an alternative form, obtained by rearranging the pseudovector-coupling external-nucleon-line-insertion terms which appear there into pseudoscalar-coupling Born terms of the usual form. This gives

$$\langle N(p_2)\,\pi(q)|\mathcal{J}(0)|N(p_1)\rangle = -\mathfrak{N}_{N\pi}\bar{u}(p_2)\Big[\frac{g_r}{M_N g_A}J_j'(k-q) + \frac{g_r}{2M_N}\{\gamma_5\tau_j, J(k)\}_+$$
$$+\frac{g_r}{2M_N}\gamma_5\tau_j\frac{\not p_2 + \not q + M_N}{\nu - \nu_B}J(k) - J(k)\frac{\not p_1 - \not q + M_N}{\nu + \nu_B}\frac{g_r}{2M_N}\gamma_5\tau_j$$
$$+\text{possible additional pion-pole “seagull” contribution}\Big]u(p_1)\,\psi_j^* + O(q) , \tag{40}$$

with

$$\nu = (p_1 + p_2)\cdot k/(2M_N) ,$$
$$\nu_B = -q\cdot k/(2M_N) , \tag{41}$$

and with all other quantities as defined above. The anticommutator term which has appeared in Eq. (40) is the PCAC (partial conservation of axial-vector current) “consistency-condition” term[25] arising from the pseudovector-to-pseudoscalar rearrangement.

With the aid of Eq. (4), we can now proceed to discuss the pion-production model, which is an extension of a calculation of weak pion production in the (3, 3) resonance region which we have described in detail elsewhere.[6] In its original form, the model included the pseudoscalar-coupling Born terms and the pion-pole terms of Eq. (40), with no kinematic approximations. In addition, the dominant (3, 3) multipoles were unitarized by the method used in the CGLN treatment of pion photoproduction,[26] so as to correctly describe (3, 3) resonance excitation. Our basic extended pion-production model is obtained by adding the commutator term in Eq. (40) (evaluated at $q = 0$, except where a pion pole appears) and the “consistency-condition” term to the Born approximation and resonant terms of the original model, yielding a pion-production amplitude which has the correct soft-pion limit. In terms of the amplitudes $V_j^{(\pm 0)}$, $j = 1, \ldots, 6$ and $A_j^{(\pm)}$, $j = 1, \ldots, 8$ used in Ref. 6 the additions are[27]

$$\Delta V_1^{(+)} \approx \frac{g_r}{M_N} F_2^V(k^2) ,$$
$$\Delta V_1^{(0)} \approx \frac{g_r}{M_N} F_2^S(k^2) ,$$
$$\Delta V_6^{(-)} \approx \frac{g_r}{M_N} \left[\frac{g_A(k^2)}{g_A} - F_1^V(k^2)\right](k^2)^{-1} ,$$
$$\Delta A_2^{(-)} \approx \frac{g_r}{M_N g_A} F_2^V(k^2) , \tag{42}$$
$$\Delta A_4^{(-)} \approx -\frac{g_r}{M_N^2 g_A} [F_1^V(k^2) - g_A g_A(k^2) + 2M_N F_2^V(k^2)] .$$

Note that the terms referred to in Ref. 6 as "dispersion-relation corrections to the small partial waves" are omitted from the amplitude, since including them along with the additions of Eq. (42) would involve double counting (and also for the practical reason that the numerical evaluation of the dispersion-relation terms is very costly in terms of computer time).

A further elaboration on the pion-production model consists of adding in the leading corrections (of first order in the pion four-momentum q and zeroth order in the lepton four-momentum transfer k) to the soft-pion limit. These corrections are calculated by the method of Low[26] and Adler and Dothan[26]; for the vector amplitude they vanish (as a result of vector current conservation), while for the isovector axial-vector amplitude they are related by PCAC to momentum derivatives of the pion-nucleon scattering amplitude at the crossing symmetric point. For an isoscalar axial-vector current the $O(q)$ corrections cannot be precisely calculated, but an heuristic resonance dominance argument given in Appendix B suggests that they may be relatively considerably smaller than in the isovector axial-vector case, and so we neglect them. For the isovector axial-vector amplitudes, the calculations of Ref. 28 tell us that[27]

$$[A_1^{(-)} - A_1^{(-)B}]\,|_0 \approx \frac{g_A}{g_r} \bar{B}^{\pi N(-)}\Big|_{\nu=\nu_B=0} - \frac{g_r}{2M_N^2}\left(\frac{1}{g_A} - g_A + \frac{\mu^V}{g_A}\right) \approx 0.36 ,$$
$$[A_3^{(+)} - A_3^{(+)B}]\,|_0 \approx \frac{g_A}{g_r} \frac{\partial \bar{A}^{\pi N(+)}}{\partial \nu_B}\Big|_{\nu=\nu_B=0} \approx 2.8 ; \tag{43a}$$
$$|_0 \equiv |_{\nu=\nu_B=k^2=0} ,$$

with the superscript B indicating the Born approximation and with the numerical values in units in which $M_\pi = 1$. In order to apply Eq. (43a), we must first estimate the extent to which the amplitudes $A_1^{(-)}, A_3^{(+)}$ in our basic pion-production model differ from their Born approximations as a result of unitarization of the (3, 3) multipoles. This is done at the end of Appendix B, with the result

$$[A_1^{(-)} - A_1^{(-)B}]\,|_0^{\text{basic model}} \approx 0.21 ,$$
$$[A_3^{(+)} - A_3^{(+)B}]\,|_0^{\text{basic model}} \approx 0.84 . \tag{43b}$$

Hence to bring the basic model into agreement with Eq. (43a) we add the $O(q)$ corrections

$$\Delta A_1^{(-)} \approx 0.15\,(1 + k^2/M^2)^{-2} ,$$
$$\Delta A_3^{(+)} \approx 1.96\,(1 + k^2/M^2)^{-2} . \tag{43c}$$

Only the $k^2 = 0$ values of the correction terms are actually determined by low-energy theorem arguments; however, to avoid a spurious dominance of these correction terms at large k^2, we have included an *ad hoc* dipole form factor[27] $(1 + k^2/M^2)^{-2}$, characterized by a dipole mass M. In the numerical work of Sec. IV, M was taken equal to the nucleon mass $M_N = 0.94$ GeV, which is rather typical of the dipole mass values[29] found in both the vector and the axial-vector form factors. As we will see in Sec. IV A, substantial variations of M about this value have a relatively small effect on the magnitude of the $O(q)$ corrections to the threshold pion-production cross sections. While the inclusion of the order-q corrections may be an improvement in the amplitude near threshold (or at a minimum, should give an idea of the likely importance of corrections to the basic soft-pion matrix element), their undamped growth as q increases makes their inclusion of doubtful value away from the threshold region. To illustrate this, we also evaluate the $O(q)$ corrections according to the modified recipe

$$\Delta A_1^{(-)} \approx (M_N/W)\,0.15\,(1 + k^2/M^2)^{-2} ,$$
$$\Delta A_3^{(+)} \approx (M_N/W)\,1.96\,(1 + k^2/M^2)^{-2} , \tag{43d}$$

which agrees with Eq. (43c) at $\nu = \nu_B = 0$, but which grows less rapidly with increasing W. To sum up, the fully extended pion-production model which we have just described contains both nucleon and pion Born diagrams with no kinematic approximations, includes the dominant (3, 3) multipoles in unitarized form, and agrees with all low-energy-theorem constraints through terms of first order in q and k, with an error of order qk at most. It should thus give a reasonably accurate description of low-invariant-maaa pion production, particularly in the region close to invariant-mass threshold.

IV. NUMERICAL RESULTS

We turn now to numerical calculations using the pion-production model developed above. In Sec. IV A we give the results of numerical studies of the model, in which we examine the effect of the $O(q)$ corrections of Eqs. (43c), (43d) and explore their sensitivity to the mass parameter M, and in which we compare the predictions of the model for charged-current-induced neutrino pion production with experiment. In Sec. IV B we use the model to give bounds on the ANL neutral-current-induced threshold pion-production cross section, for a variety of different models for the structure of the weak neutral current. Finally, in Sec. IV C we study low-invariant-mass pion production in the BNL neutrino flux, in the special case of a pure isoscalar weak neutral current.

A. Numerical studies of the model

We begin our numerical examination of the pion-production model of Sec. III C with a study of the $O(q)$ correction terms added to the weak pion production amplitude in Eq. (43). In Table II we give theoretical ANL 2-bin cross sections, defined as in Eq. (16), for the seven allowed charged- and neutral-current-induced pion-production reactions. In column 2 we give the cross section obtained without the $O(q)$ correction [that is, from the basic pion-production model including the soft-pion additions of Eq. (42), but without the additions of Eqs. (43c) or (43d)]. In columns 3, 4, and 5 we give the corresponding cross sections with the $O(q)$ corrections included as in Eq. (43c), taking the *ad hoc* dipole mass M as M_N, $M_N/\sqrt{2}$, and $M_N\sqrt{2}$, respectively. In column 6 we give the cross sections calculated with the $O(q)$ corrections included as in Eq. (43d), with $M=M_N$. We see that the $O(q)$ corrections have a substantial effect on threshold cross sections for $\nu+p\to\nu+p+\pi^0$ and $\nu+n\to\nu+n+\pi^0$, a moderate effect on the cross section for $\nu+p\to\mu^-+p+\pi^+$, and a relatively small effect on the remaining reactions. As expected in the threshold region, the recipes of Eq. (43c) and Eq. (43d) for the $O(q)$ corrections give similar results; we also see that the variation in the threshold cross section as M^2 is changed by a factor of 2 in either direction from $M^2=M_N{}^2$ is smaller than the effect of including the $O(q)$ corrections, indicating that the sensitivity to the value of the mass parameter M is not excessive. In Table III we show the effect of the $O(q)$ additions on the ANL and BNL cross sections integrated over the low-invariant-mass region $W\le 1.4$ GeV. Again, the reactions $\nu+N\to\nu+N+\pi^0$ are sensitive to the $O(q)$ additions, with the effects on the other cross sections ranging from moderate to small. Here, however, we see a substantial dependence on whether the recipe of Eq. (43c) or of Eq. (43d) is used, indicating that the $O(q)$ additions do not constitute a well-defined correction to the basic pion production model outside the threshold region. A satisfactory treatment of the $O(q)$ terms away from threshold would require their interpretation as the low-energy limits of appropriate particle exchange terms. In the numerical work on the ANL threshold cross sections for $\nu+n\to\nu+p+\pi^-$ described in Sec. IV B, we will include the $O(q)$ correction terms with $M=M_N$. In the numerical work of Sec. IV C analyzing the isoscalar case for the BNL spectrum, we will neglect the $O(q)$ corrections—they vanish for an isoscalar vector current and, as argued in Appendix B, may be relatively small (although hard to estimate precisely) for an isoscalar axial-vector current.

Obviously, the best way to assess the reliability of the pion-production model developed in Sec. III C is to compare its predictions for charged-current-induced pion production reactions with experiment. In Fig. 1 the ANL data[30] for $\nu+p\to\mu^-+p+\pi^+$ are plotted together with predictions

TABLE II. Effect of $O(q)$ additions of Eqs. (43c), (43d) and sensitivity to the *ad hoc* dipole parameter M in the ANL threshold region. Neutral-current cross sections are calculated in the Weinberg-Salam model, with $\sin^2\theta_W=0.35$ and $\Delta\mathcal{J}^\lambda=0$ [see Eq. (52)].

Reaction	Values of $\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}}$ in 10^{-41} cm^2: Without $O(q)$	With $O(q)$ from Eq. (43c): $M=M_N$	$M=M_N/\sqrt{2}$	$M=M_N\sqrt{2}$	With $O(q)$ from Eq. (43d)
$\nu+n\to\mu^-+p+\pi^0$	3.8	3.5	3.6	3.5	3.6
$\nu+p\to\mu^-+p+\pi^+$	4.5	5.9	5.6	6.4	5.7
$\nu+n\to\mu^-+n+\pi^+$	2.3	2.1	2.0	2.2	2.0
$\nu+p\to\nu+p+\pi^0$	0.42	0.73	0.61	0.88	0.64
$\nu+n\to\nu+n+\pi^0$	0.47	0.77	0.65	0.92	0.69
$\nu+n\to\nu+p+\pi^-$	0.91	0.80	0.82	0.77	0.81
$\nu+p\to\nu+n+\pi^+$	0.97	0.87	0.89	0.84	0.88

TABLE III. Effect of $O(q)$ additions of Eqs. (43c), (43d) in the ANL and BNL (3, 3) resonance regions, defined by $W \leq 1.4$ GeV. Neutral-current cross sections are calculated in the Weinberg-Salam model, with $\sin^2\theta_W = 0.35$ and $\Delta\mathcal{J}^\lambda = 0$ [see Eq. (52)]. As is evident from the differences between the two recipes for adding in the $O(q)$ terms away from the threshold region, the $O(q)$ additions do not constitute a well-defined correction to the basic pion production model when the entire (3, 3) resonance region is considered. However, they do usefully indicate which channels may prove to be particularly sensitive to corrections to the basic model.

Reaction	Values of σ^{ANL} ($W \leq 1.4$ GeV) in 10^{-38} cm^2: Without $O(q)$	With $O(q)$ Eq. (43c)	With $O(q)$ Eq. (43d)	Values of σ^{BNL} ($W \leq 1.4$ GeV) in 10^{-38} cm^2: Without $O(q)$	With $O(q)$ Eq. (43c)	With $O(q)$ Eq. (43d)
$\nu + n \to \mu^- + p + \pi^0$	0.0733	0.0694	0.0698	0.147	0.143	0.143
$\nu + p \to \mu^- + p + \pi^+$	0.219	0.237	0.228	0.427	0.499	0.466
$\nu + n \to \mu^- + n + \pi^+$	0.0571	0.0534	0.0491	0.129	0.134	0.116
$\nu + p \to \nu + p + \pi^0$	0.0283	0.0348	0.0308	0.0532	0.0765	0.0629
$\nu + n \to \nu + n + \pi^0$	0.0288	0.0350	0.0311	0.0539	0.0768	0.0634
$\nu + n \to \nu + p + \pi^-$	0.0192	0.0181	0.0181	0.0366	0.0352	0.0351
$\nu + p \to \nu + n + \pi^+$	0.0204	0.0191	0.0192	0.0383	0.0367	0.0367

of the pion production model, both with $O(q)$ additions (curves b and c) and without these additions (curve a). The theoretical curves are evidently low by 30–40% in the case of curve a and by smaller amounts in the cases of curves b and c. Part of this discrepancy may arise from uncertainties in the absolute level of the ANL neutrino flux (these uncertainties are included in the experimental error bars) and in the value of the axial-vector mass parameter[31] M_A, but part is probably due to the known[6] tendency of the pion-production model to underestimate pion-production cross sections for $|k^2| \geq 0.6$ (GeV/c)^2. To minimize this problem, in discussing the BNL isoscalar-current case in Sec. IV C we will always compare *ratios* of cross sections computed within the pion production model with the corresponding ratios obtained experimentally, rather than making direct comparisons of cross sections between theory and experiment. In Table IV we compare preliminary ANL values[32] of the ratios $\sigma(\nu+n \to \mu^- + p + \pi^0)/\sigma(\nu+p \to \mu^- + p + \pi^+)$ and $\sigma(\nu+n \to \mu^- + n + \pi^+)/\sigma(\nu+p \to \mu^- + p + \pi^+)$ with the corresponding theoretical predictions for the invariant-mass interval $W \leq 1.4$ GeV. The agreement is seen to be generally satisfactory. In Fig. 2 we compare the area normalized theoretical invariant-mass distribution for $\nu + p \to \mu^- + p + \pi^+$ [including $O(q)$ corrections from Eq. (43c)] with the corresponding ANL experimental histogram[33]; the agreement in this case is excellent. In Fig. 3 we give the same comparison for the reactions[34] $\nu + n \to \mu^- + p + \pi^0$ and $\nu + n \to \mu^- + n + \pi^+$. The agreement is again satisfactory. In general, the comparisons given above suggest that the pion-production model developed in Sec. III C should be reliable to better than a factor of 2 in the region at and below resonance. The reliability should be substantially better than this for relative cross-section ratios or reactions without large $O(q)$ corrections.

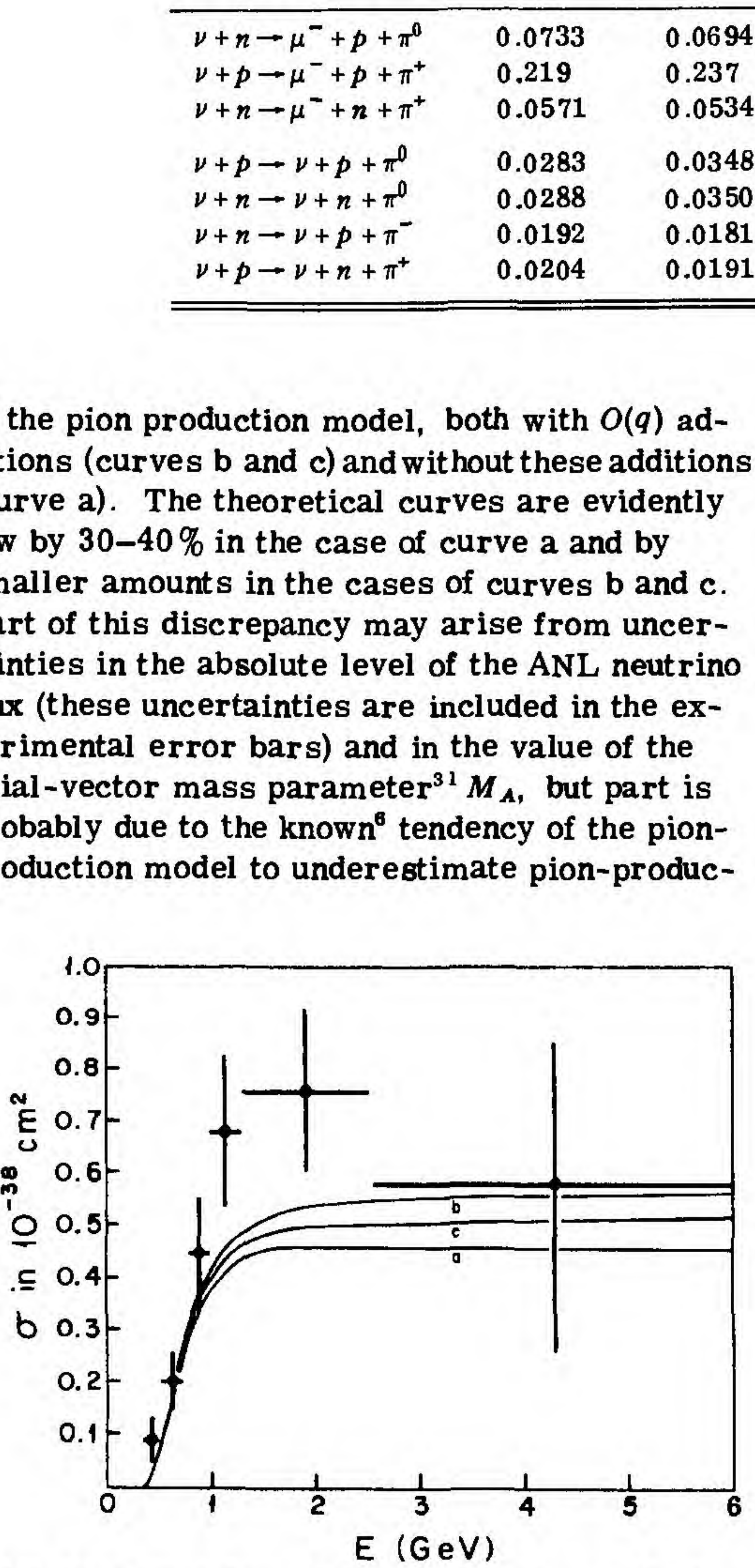


FIG. 1. Comparison of the extended pion-production model of Sec. III C with the ANL data for $\nu + p \to \mu^- + p + \pi^+$. Curve a, basic model containing Born, resonant, and soft-pion terms; curve b, basic model with $O(q)$ additions from Eq. (43c); curve c, basic model with $O(q)$ additions from Eq. (43d).

B. Threshold neutral-current-induced pion production in the ANL flux

We consider now the application of the formulas developed in Sec. III to the study of neutral-cur-

TABLE IV. Comparison of theoretical predictions for charged-current pion final-state ratios with preliminary ANL experimental results (Ref. 32).

Ratio	Experiment	Theory: Without $O(q)$	Theory: With $O(q)$, Eq. (43c)	Theory: With $O(q)$, Eq. (43d)
$\dfrac{\sigma(\nu+n\to\mu^-+p+\pi^0)}{\sigma(\nu+p\to\mu^-+p+\pi^+)}$	0.27 ± 0.06	0.34	0.29	0.31
$\dfrac{\sigma(\nu+n\to\mu^-+n+\pi^+)}{\sigma(\nu+p\to\mu^-+p+\pi^+)}$	0.31 ± 0.07	0.26	0.23	0.22

rent-induced threshold pion production in the ANL neutrino flux.[4] We start with the general six-parameter neutral-current structure in Eq. (18), but we note that since the isoscalar axial currents contribute only[35] through the $g_A^{(0,8)}\gamma^\lambda\gamma_5$ term in their nucleon vertices, the effective number of parameters entering the pion production calculation can be reduced to 5. These are conveniently introduced by writing the one-nucleon matrix element of the neutral current as

$$\langle N(p_2)|\mathcal{J}_N^\lambda|N(p_1)\rangle=\mathfrak{N}_N\bar{u}(p_2)\{[-\lambda_1 g_A(k^2)\gamma^\lambda\gamma_5+\lambda_2F_1^V(k^2)\gamma^\lambda+i\lambda_2F_2^V(k^2)\sigma^{\lambda\eta}k_\eta]\tfrac{1}{2}\tau_3$$
$$+[-\lambda_3D(k^2)\gamma^\lambda\gamma_5+\lambda_4D(k^2)\gamma^\lambda+i\lambda_5(2M_N)^{-1}D(k^2)\sigma^{\lambda\eta}k_\eta]\tfrac{1}{2}\}u(p_1)\,, \tag{44}$$

with $D(k^2)$ a dipole structure characterizing the isoscalar current vertices, which for definiteness we take as

$$D(k^2)=(1-k^2/M_N{}^2)^{-2}\,. \tag{45}$$

To a first approximation, we expect[36] that small changes in the isoscalar dipole mass parameter from the assigned value of M_N can be compensated by making appropriate rescalings of the isoscalar parameters $\lambda_{3,4,5}$. In terms of the parameters $\lambda_1,\ldots,\lambda_4$, the couplings $g_{V0,3,8}, g_{A0,3,8}$ introduced in Eq. (18) are given by

$$g_{V0}(\tfrac{2}{3})^{1/2}+g_{V8}(\tfrac{1}{3})^{1/2}=\tfrac{1}{3}\lambda_4,\quad g_{V3}=\lambda_2\,,$$
$$g_{A0}(\tfrac{2}{3})^{1/2}+g_{A8}(\tfrac{1}{3})^{1/2}=\frac{1}{g_A^S}\lambda_3,\quad g_{A3}=\lambda_1\,, \tag{46a}$$

with g_A^S an effective isoscalar axial-vector renormalization constant defined by

$$g_A^S=\frac{g_{A0}(\frac{2}{3})^{1/2}g_A^{(0)}(0)+g_{A8}(\frac{1}{3})^{1/2}g_A^{(8)}(0)}{g_{A0}(\frac{2}{3})^{1/2}+g_{A8}(\frac{1}{3})^{1/2}}\,. \tag{46b}$$

In terms of these definitions, the deep-inelastic constraint of Eq. (36) becomes

$$1.5\geq 3R_\nu+R_{\bar\nu}\geq\lambda_1{}^2+\lambda_2{}^2, \tag{47}$$

while the stronger constraint of Eq. (37), which follows when quark-parton-model information is used, takes the form

$$1.5\geq 3R_\nu+R_{\bar\nu}=\lambda_1{}^2+\lambda_2{}^2+\frac{\lambda_3{}^2}{(g_A^S)^2}+\tfrac{1}{9}\lambda_4{}^2. \tag{48}$$

Equations (15) and (47), or (15) and (48), are the basic constraints which will be imposed in maximizing $\sigma_{2\text{bin}}^{\text{ANL}}$ over the space of parameters $\lambda_1,\ldots,\lambda_5$.

Obviously, to recompute the pion-production and neutrino-proton-elastic-scattering cross sections for each distinct set of parameter values being studied would be a very inefficient procedure from

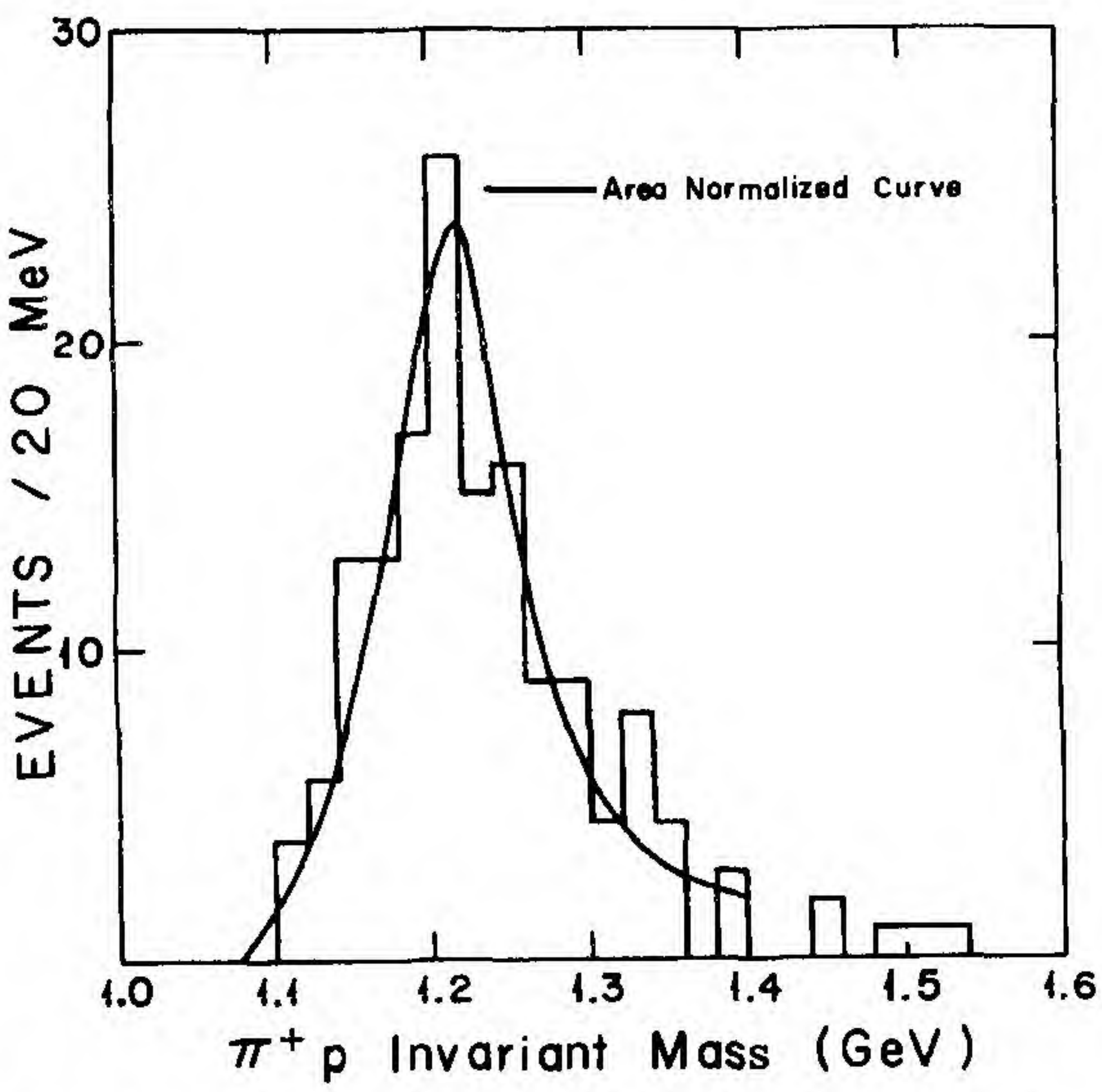


FIG. 2. Comparison of the area-normalized theoretical invariant-mass distribution for $\nu+p\to\mu^-+p+\pi^+$ [calculated with $O(q)$ additions from Eq. (43c)] with the ANL histogram for this reaction.

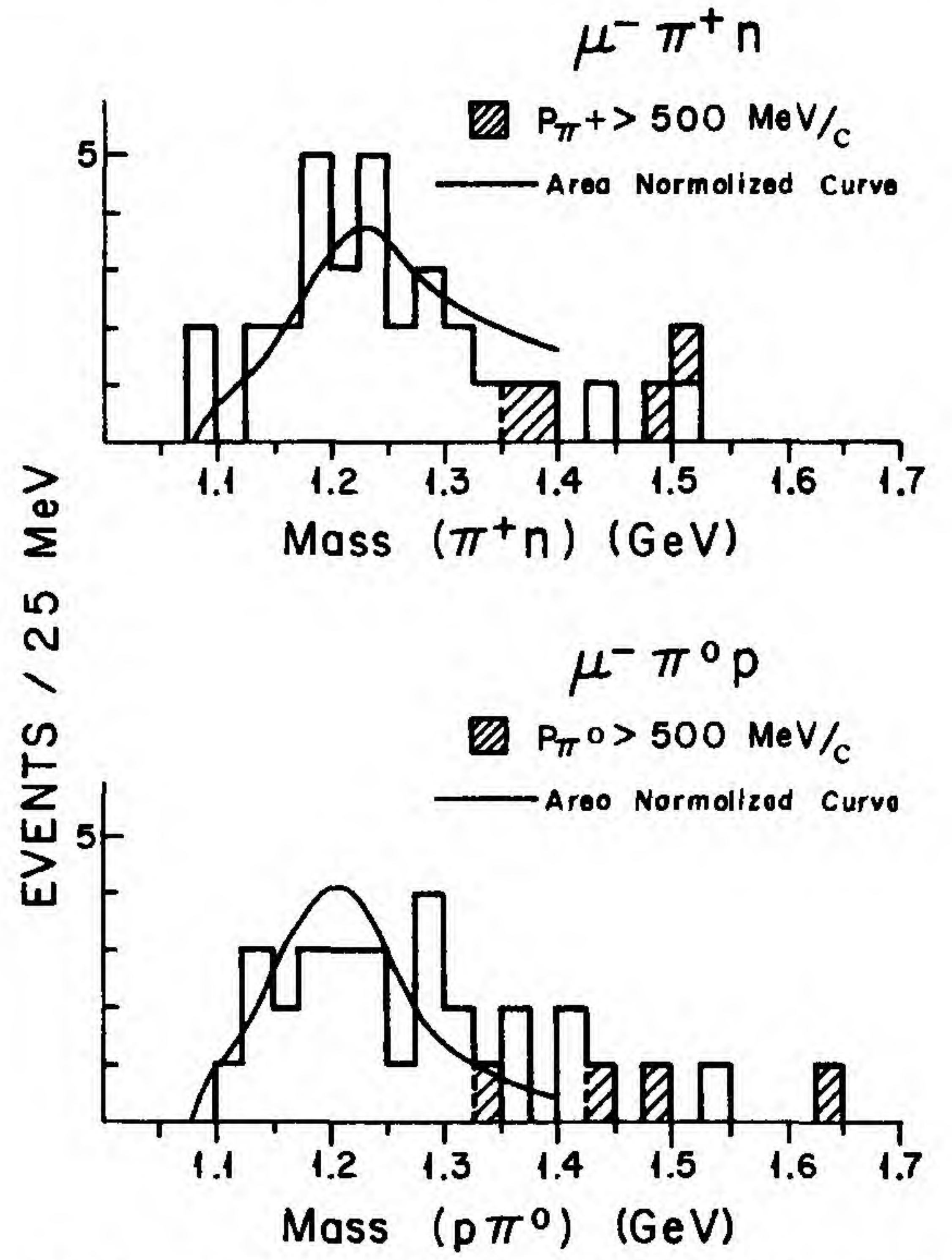


FIG. 3. Comparison of the area-normalized theoretical invariant-mass distributions for $\nu+n\to\mu^-+n+\pi^+$ and $\nu+n\to\mu^-+p+\pi^0$ [calculated with $O(q)$ additions from Eq. (43c)] with the corresponding ANL histograms. The theoretical predictions have been folded with the experimental invariant-mass resolutions of 25 MeV for $n+\pi^+$ and 40 MeV for $p+\pi^0$.

a numerical point of view. Rather, we exploit the fact that the cross sections are quadratic forms in the parameters λ_j, taking the form

$$\begin{aligned}\sigma_{2\,\text{bin}}^{\text{ANL}}/10^{-38}\ \text{cm}^2 &= \sum_{l=1}^{5}\sum_{1\le j\le l} P_{ij}\lambda_i\lambda_j,\\ \sigma^{\text{ANL}}(\nu+p\to\nu+p)/10^{-38}\ \text{cm}^2 &= \sum_{l=1}^{5}\sum_{1\le j\le l} E_{ij}\lambda_i\lambda_j,\end{aligned}\tag{49}$$

so it is only necessary to perform the cross-section calculation for the 15 parameter sets

$$\left.\begin{aligned}&\lambda_i=0,\quad i\ne I,J\\ &\lambda_I=1,\quad \lambda_J=1\end{aligned}\right\}\ 1\le I\le J\le 5\tag{50}$$

to extract the coefficients[4] P_{ij}, E_{ij}, which are tabulated in Table V. The quadratic forms of Eq. (49) are then used to compute the cross sections when searching over parameter values, permitting a complete survey of the five-parameter space using a very reasonable amount of computer time.

TABLE V. Coefficients determining $\sigma_{2\,\text{bin}}^{\text{ANL}}$ and $\sigma^{\text{ANL}}(\nu+p\to\nu+p)$ via the quadratic forms of Eq. (49). Note the comment in Ref. 4.

Pion-production coefficients		Elastic-scattering coefficients	
P_{11}	0.621×10^{-3}	E_{11}	0.692×10^{-1}
P_{22}	0.807×10^{-3}	E_{22}	0.767×10^{-1}
P_{33}	0.163×10^{-3}	E_{33}	0.478×10^{-1}
P_{44}	0.244×10^{-4}	E_{44}	0.364×10^{-1}
P_{55}	0.121×10^{-4}	E_{55}	0.300×10^{-2}
P_{12}	0.534×10^{-3}	E_{12}	0.656×10^{-1}
P_{13}	0.772×10^{-5}	E_{13}	0.115
P_{14}	0.166×10^{-4}	E_{14}	0.158×10^{-1}
P_{15}	-0.211×10^{-4}	E_{15}	0.159×10^{-1}
P_{23}	-0.393×10^{-4}	E_{23}	0.554×10^{-1}
P_{24}	0.143×10^{-3}	E_{24}	0.858×10^{-1}
P_{25}	-0.312×10^{-4}	E_{25}	0.201×10^{-1}
P_{34}	0.328×10^{-4}	E_{34}	0.134×10^{-1}
P_{35}	0.532×10^{-4}	E_{35}	0.134×10^{-1}
P_{45}	0.996×10^{-5}	E_{45}	0.229×10^{-2}

The results, for various assumptions about the structure of the weak neutral current, are as follows:

(1) Pure isoscalar weak neutral current. Taking $\lambda_1=\lambda_2=0$ and maximizing $\sigma_{2\,\text{bin}}^{\text{ANL}}$ over the $\lambda_3,\lambda_4,\lambda_5$ subspace subject to the constraint of Eq. (15) gives the upper bound

$$\sigma_{2\,\text{bin}}^{\text{ANL}}\le 1.0\times10^{-41}\ \text{cm}^2.\tag{51}$$

(2) Weinberg-Salam SU(2)⊗U(1) model. In the simplest, one-parameter version of this model,[37] the neutral current has the form

$$\begin{aligned}\mathcal{J}_N^\lambda &= \mathfrak{F}_3^\lambda-\mathfrak{F}_3^{5\lambda}-2x(\mathfrak{F}_3^\lambda+3^{-1/2}\mathfrak{F}_8^\lambda)+\Delta\mathcal{J}^\lambda,\\ x &= \sin^2\theta_W,\end{aligned}\tag{52}$$

with $\Delta\mathcal{J}^\lambda$ an isoscalar $V-A$ strangeness and "charm" current contribution which is conventionally assumed to couple only weakly to nonstrange low-mass hadrons (such as the low-energy pion-nucleon system). Neglecting $\Delta\mathcal{J}^\lambda$ for the moment we can make an absolute calculation of the cross section for $\nu+n\to\nu+p+\pi^-$, giving

$$\sigma_{2\,\text{bin}}^{\text{ANL}}=0.75\times10^{-41}\ \text{cm}^2.\tag{53}$$

In certain extensions of the original Weinberg-Salam model, the neutral current has the general form of Eq. (52), but with an adjustable strength parameter κ in front,

$$\mathcal{J}_N^\lambda=\kappa[\mathfrak{F}_3^\lambda-\mathfrak{F}_3^{5\lambda}-2x(\mathfrak{F}_3^\lambda+3^{-1/2}\mathfrak{F}_8^\lambda)]+\Delta\mathcal{J}^\lambda.\tag{54}$$

Still neglecting $\Delta\mathcal{J}^\lambda$, and comparing with Eq. (44), we now see that the parameters λ_j have the values

$$\lambda_1 = \kappa,$$
$$\lambda_2 = \kappa(1-2x), \quad \lambda_4 = -2\kappa x, \tag{55}$$
$$\lambda_3 = 0. \qquad \lambda_5 = 0.24\kappa x.$$

Maximizing $\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}}$ over the κ, x parameter space (allowing all real values of x, rather than restricting x to lie between 0 and 1) subject to the constraints of Eqs. (15) and (47) gives the upper bound

$$\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}} \leq 1.5\times 10^{-41}\ \mathrm{cm}^2. \tag{56}$$

Finally, we can include the isoscalar addition $\Delta\mathcal{J}^\lambda$ by regarding $\lambda_3, \lambda_4, \lambda_5$ as free parameters, rather than relating them to κ and x as in Eq. (55). Searching now over the five-parameter $\kappa, x, \lambda_3, \lambda_4, \lambda_5$ space (again allowing all real values of x) subject to the constraints of Eqs. (15) and (47) gives the upper bound

$$\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}} \leq 4.6\times 10^{-41}\ \mathrm{cm}^2. \tag{57}$$

We emphasize that Eq. (57) *is the upper bound on* $\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}}$ *for the most general hadronic neutral current formed from the usual vector and axial-vector nonets*. If Eq. (47) is replaced by the stronger constraint of Eq. (48), and if the parameters $g \equiv \lambda_5/\lambda_4$ and g_A^S are restricted [as suggested by the quark-model[16,24] values of Eq. (39)] by

$$|g| \leq 1.5, \quad |g_A^S| \leq 0.74, \tag{58}$$

then the bound in the general V, A case is substantially reduced, to

$$\sigma_{2\,\mathrm{bin}}^{\mathrm{ANL}} \leq 1.5\times 10^{-41}\ \mathrm{cm}^2. \tag{59}$$

C. Analysis of low-invariant-mass ($W \leq 1.4$ GeV) pion production at BNL: Isoscalar current case

We turn finally to an analysis of low-invariant-mass ($W \leq 1.4$ GeV) pion production in the BNL neutrino flux.[38] Recently, the Columbia-Illinois-Rockefeller collaboration at BNL has reported a measurement of the ratio

$$R_0' = \frac{\sigma(\nu + T \to \nu + \pi^0 + \cdots)}{2\sigma(\nu + T \to \mu^- + \pi^0 + \cdots)}, \qquad T = \tfrac{1}{4}({}_6\mathrm{C}^{12}) + \tfrac{3}{4}({}_{13}\mathrm{Al}^{27}), \tag{60}$$

with the preliminary result[3,39]

$$R_0' = 0.17 \pm 0.06. \tag{61}$$

This measured value of R_0' is in accord with the value expected[40] in the Weinberg-Salam model when $\sin^2\theta_W$ is in the currently favored range of 0.3–0.4. Hence *if* (3, 3) resonance excitation, which is expected in the Weinberg-Salam model (see Fig. 4), is observed in the BNL experiment, the presumption would be strongly in favor of the standard gauge-theory interpretation of neutral currents. However, preliminary BNL invariant-mass spectra[3] for π^0 production in the charged- and neutral-current cases show a clear (3, 3) peak in the charged-current case, but indicate no comparable peaking in the neutral-current reaction, suggesting that perhaps the (3, 3) resonance is not excited by the neutral current. In what follows we analyze the implications for neutral-current structure if this indication is confirmed both by more detailed analysis of the BNL data and by other experiments.

Since the isovector V and A neutral currents both[41] strongly excite the (3, 3) resonance, the absence of a (3, 3) peak in the V, A case would suggest an isoscalar neutral-current structure, and we assume this in what follows. Applying nuclear charge-exchange corrections as described in Appendix C, we find that the nuclear target ratio quoted in Eq. (61) implies the free-nucleon target ratio

$$2R_0 \equiv \frac{\sigma^{\mathrm{BNL}}(\nu + n \to \nu + n + \pi^0) + \sigma^{\mathrm{BNL}}(\nu + p \to \nu + p + \pi^0)}{\sigma^{\mathrm{BNL}}(\nu + n \to \mu^- + p + \pi^0)}$$
$$= 2R_0' \times 1.4 = 0.48 \pm 0.17. \tag{62}$$

Let us now compare the experimental result of Eq. (62) with theoretical predictions obtained from the extended pion production model developed in

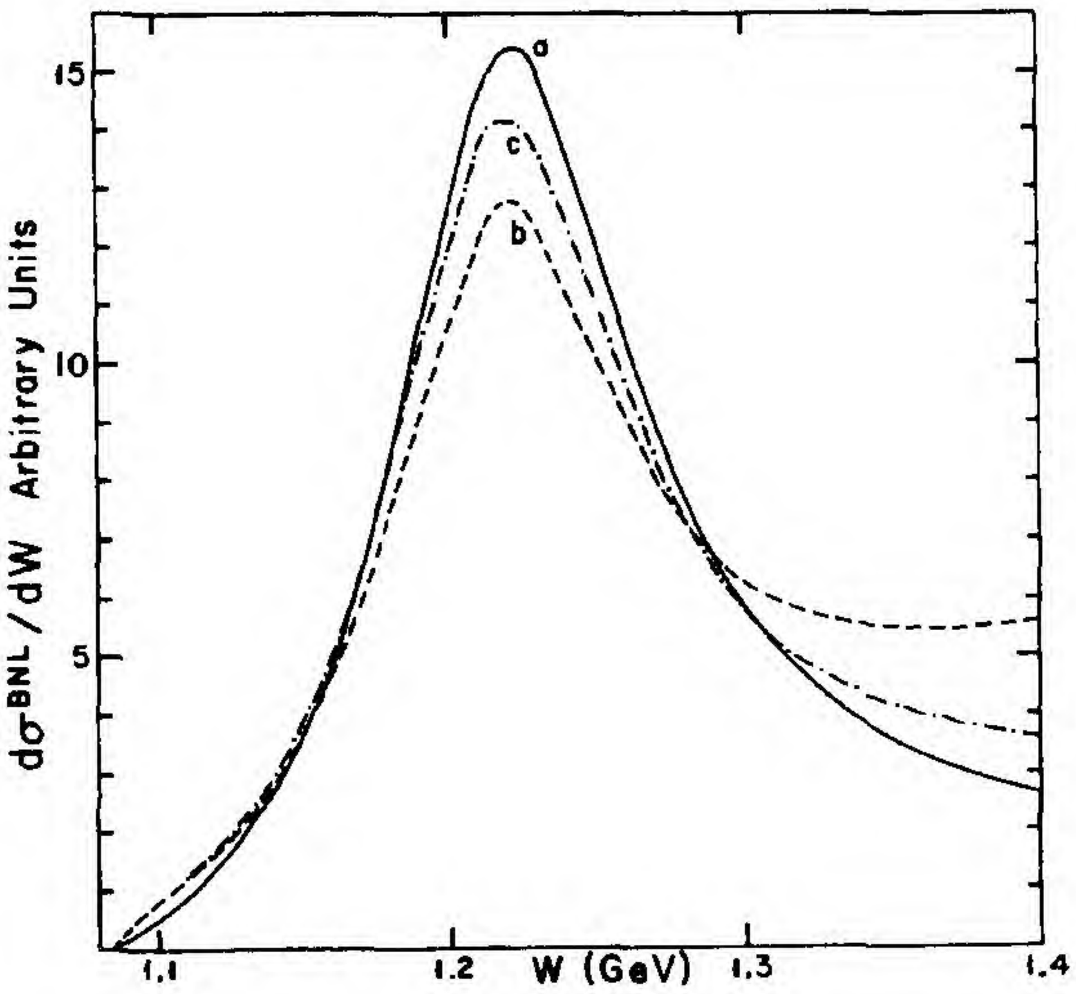


FIG. 4. Area-normalized, BNL-flux-averaged theoretical invariant-mass distribution in the Weinberg-Salam model (with $\sin^2\theta_W = 0.35$) for the reaction $\nu + p \to \nu + p + \pi^0$. Curve a, basic model containing Born, resonant, and soft-pion terms; curve b, basic model with $O(q)$ additions from Eq. (43c); curve c, basic model with $O(q)$ additions from Eq. (43d).

Sec. III C, in the case of a pure isoscalar neutral current. Again we parameterize the neutral current as in Eq. (44), with $\lambda_1=\lambda_2=0$. The BNL flux-averaged pion production and elastic cross sections are then quadratic forms in $\lambda_3, \lambda_4, \lambda_5$, taking the form

$$[\sigma^{\mathrm{BNL}}(\nu+n\to\nu+n+\pi^0,\ W\le W_M)+\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p+\pi^0,\ W\le W_M)]/10^{-38}\ \mathrm{cm}^2=\sum_{i=3}^{5}\sum_{3\le j\le i}P_{ij}(W_M)\lambda_i\lambda_j,$$

$$\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p,\ \text{Cundy cuts})/10^{-38}\ \mathrm{cm}^2=\sum_{i=3}^{5}\sum_{3\le j\le i}E_{ij}^{(1)}\lambda_i\lambda_j,$$

$$\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p,\ \text{no cuts})/10^{-38}\ \mathrm{cm}^2=\sum_{i=3}^{5}\sum_{3\le j\le i}E_{ij}^{(2)}\lambda_i\lambda_j; \tag{63}$$

Cundy cuts: $1\ \mathrm{GeV}\le E\le 4\ \mathrm{GeV},\quad 0.3\ (\mathrm{GeV}/c)^2\le |k^2|\le 1\ (\mathrm{GeV}/c)^2$.

The pion-production coefficients P_{ij} (for $W_M=1.2$, 1.3, and 1.4 GeV) and the cut and uncut[42] elastic-scattering coefficients E_{ij} required in Eq. (63) are tabulated in Table VI. In maximizing the pion-production cross section over the space of parameters $\lambda_3, \lambda_4, \lambda_5$, we impose the constraints

$$\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p,\ \text{Cundy cuts})\le 0.24\sigma^{\mathrm{BNL}}(\nu+n\to\mu^-+p,\ \text{Cundy cuts})\approx 0.085\times10^{-38}\ \mathrm{cm}^2,$$
$$1.5\ge 3R_\nu+R_{\bar\nu}=\frac{\lambda_3^2}{(g_A^S)^2}+\tfrac{1}{9}\lambda_4^2, \tag{64}$$

the first of which is the Cundy *et al.*[43] 95% confidence-level limit from the CERN neutrino experiment, which has a neutrino flux similar[43] to that of the BNL experiment, while the second is the deep-inelastic quark-parton-model constraint of Eqs. (37) and (48) above. The results of the maximization are expressed as theoretical upper bounds on the ratio $2R_0$ defined in Eq. (62), with *both* the numerator and the denominator calculated from the pion production model. As stressed in Sec. IV A, the procedure of comparing theoretical cross-section *ratios* with experimental ratios should minimize the effects of discrepancies between the experimental and theoretical cross-section magnitudes. For the denominator cross section $\sigma^{\mathrm{BNL}}(\nu+n\to\mu^-+p+\pi^0,\ W\le 1.4\ \mathrm{GeV})$ we use the value $0.143\times10^{-38}\ \mathrm{cm}^2$ listed in columns 6 and 7 of Table III, corresponding to inclusion of $O(q)$ corrections; however, as is apparent from the table, the effect of the $O(q)$ corrections on this cross section is very small.

Results of the maximization calculation[44] are given in Figs. 5 and 6. Curve a of Fig. 5 gives the maximum for an isoscalar pure vector current, while curve b gives the maximum when an octet isoscalar axial-vector current is also present (corresponding to $g_A^S=0.45$), both plotted versus the isoscalar current gyromagnetic ratio $g=\lambda_5/\lambda_4$. Evidently, both curves lie below the BNL data, with a discrepancy exceeding a factor of 2 unless $|g|\gtrsim 4$–6, that is, unless the isoscalar vector current has a $|g|$ value which is anomalously large based on quark-model expectations.[24] Interestingly, a vector current with a large $|g|$ value pro-

TABLE VI. Coefficients determining $\sigma^{\mathrm{BNL}}(\nu+n\to\nu+n+\pi^0)+\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p+\pi^0)$ for various W ranges and $\sigma^{\mathrm{BNL}}(\nu+p\to\nu+p)$, both cut and uncut, via the quadratic forms of Eq. (63).

	Pion-production coefficients			Elastic-scattering coefficients			
	$W\le1.2$ GeV	$W\le1.3$ GeV	$W\le1.4$ GeV	Cundy cuts		No cuts	
P_{33}	0.210×10^{-2}	0.607×10^{-2}	0.109×10^{-1}	$E_{33}^{(1)}$	0.176×10^{-1}	$E_{33}^{(2)}$	0.555×10^{-1}
P_{44}	0.286×10^{-3}	0.638×10^{-3}	0.994×10^{-3}	$E_{44}^{(1)}$	0.158×10^{-1}	$E_{44}^{(2)}$	0.515×10^{-1}
P_{55}	0.193×10^{-3}	0.566×10^{-3}	0.106×10^{-2}	$E_{55}^{(1)}$	0.265×10^{-2}	$E_{55}^{(2)}$	0.480×10^{-2}
P_{34}	0.247×10^{-3}	0.925×10^{-3}	0.165×10^{-2}	$E_{34}^{(1)}$	0.558×10^{-2}	$E_{34}^{(2)}$	0.105×10^{-1}
P_{35}	0.467×10^{-3}	0.138×10^{-2}	0.258×10^{-2}	$E_{35}^{(1)}$	0.558×10^{-2}	$E_{35}^{(2)}$	0.105×10^{-1}
P_{45}	0.169×10^{-3}	0.519×10^{-3}	0.941×10^{-3}	$E_{45}^{(1)}$	0.635×10^{-3}	$E_{45}^{(2)}$	0.142×10^{-2}

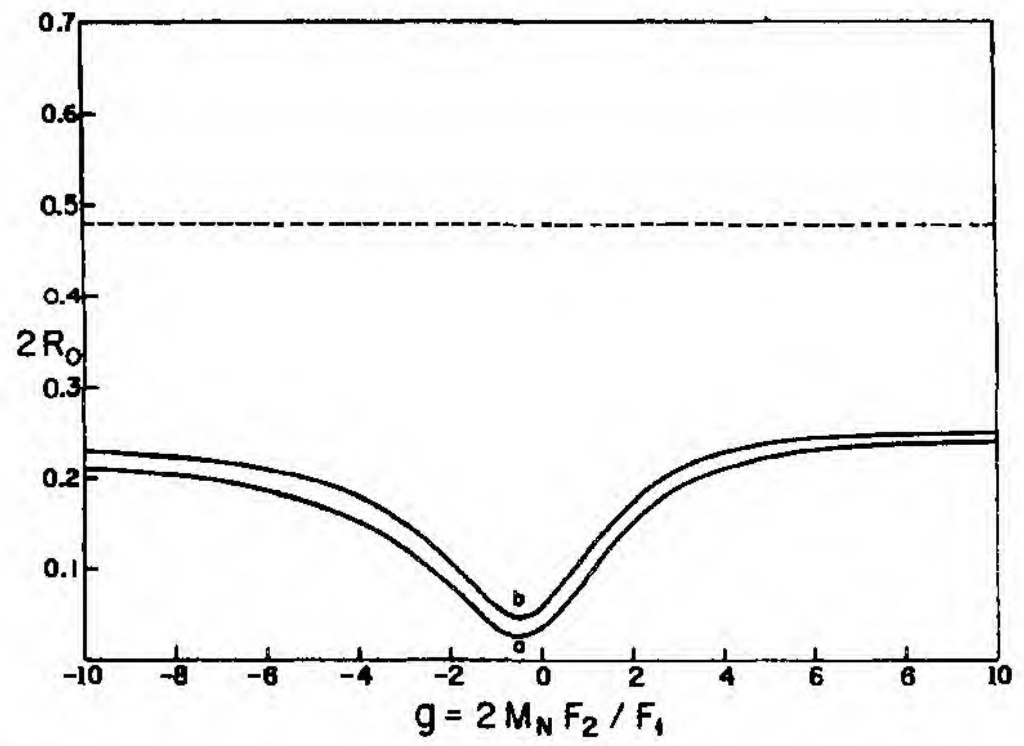


FIG. 5. Results of a maximization calculation for BNL cross-section ratios in the invariant-mass interval $W \leq 1.4$ GeV, plotted versus the g value of the isoscalar vector current. Curve a is the maximum for an isoscalar pure vector current; curve b is the maximum when an isoscalar axial current is also present, with the axial-vector renormalization constant fixed at $g_A^S = 0.45$ (the octet axial-vector current value). The dashed line is the central experimental value from Eq. (62).

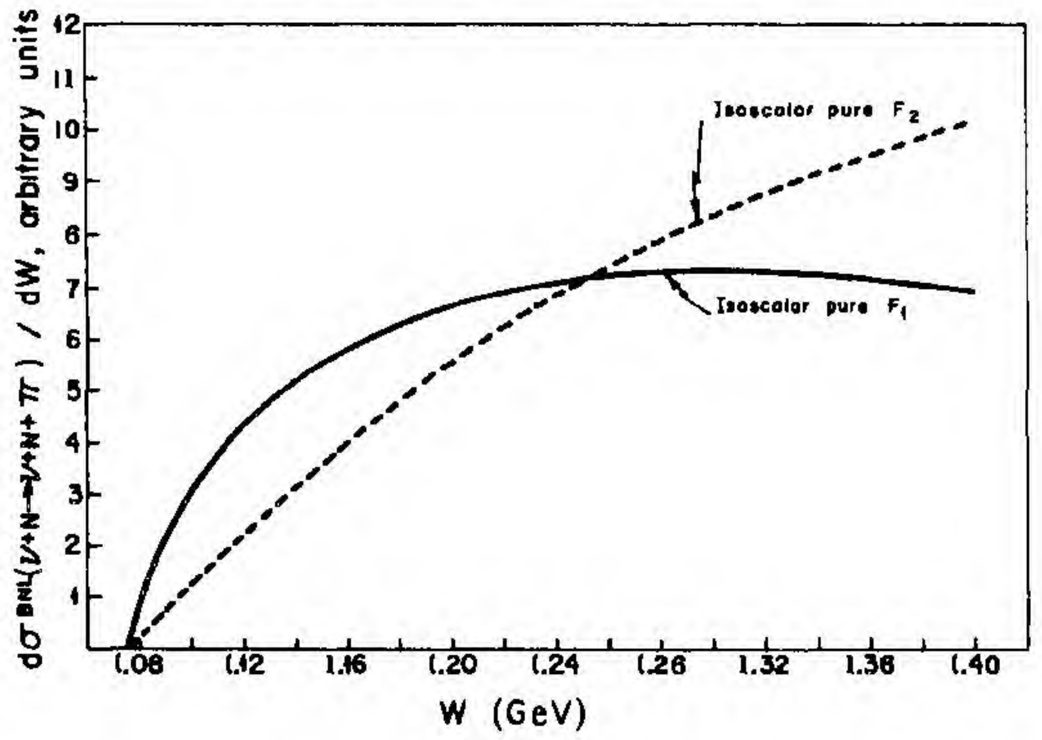


FIG. 7. Shapes of $d\sigma/dW$ for an isoscalar vector current containing an F_1 term only or an F_2 term only. The two curves are normalized to equal area for $W \leq 1.4$ GeV.

duces a characteristic change in the $d\sigma/dW$ plot predicted for the BNL flux, as shown in the dashed curve in Fig. 7. [The dashed curve is calculated for the case of an isoscalar vector current containing only an F_2 form factor; for the F_1, F_2 admixture corresponding to $|g| = 4$, the curve is substantially the same. Similarly, changing the *ad hoc* dipole mass in Eq. (45) from M_N to $M_N\sqrt{2}$ or

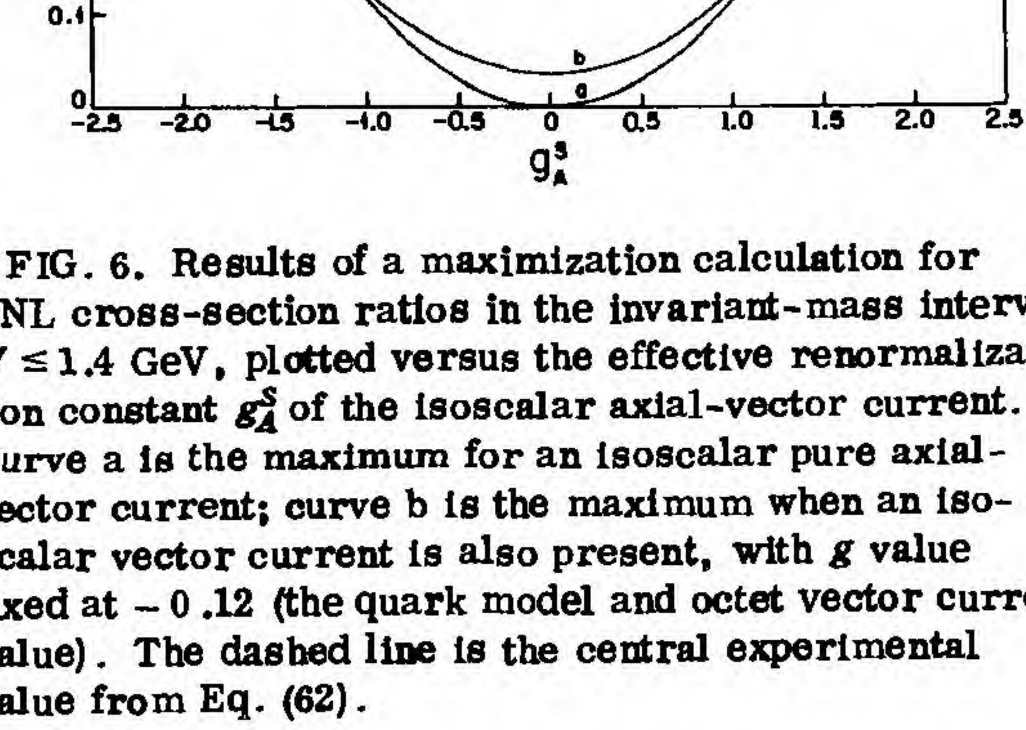


FIG. 6. Results of a maximization calculation for BNL cross-section ratios in the invariant-mass interval $W \leq 1.4$ GeV, plotted versus the effective renormalization constant g_A^S of the isoscalar axial-vector current. Curve a is the maximum for an isoscalar pure axial-vector current; curve b is the maximum when an isoscalar vector current is also present, with g value fixed at -0.12 (the quark model and octet vector current value). The dashed line is the central experimental value from Eq. (62).

$M_N/\sqrt{2}$ produces only a 2% change in the dashed curve.] As seen in the figure, an isoscalar vector current with large $|g|$ produces, relative to the pure F_1 case,[45] a depression in the $d\sigma/dW$ distribution for small W, characterized by an almost linear rise from threshold, and a corresponding enhancement at the large W end of the range. An experiment with good statistics should be able to search for this effect. Continuing with the results of the maximization calculation, curve a of Fig. 6 gives the maximum for an isoscalar pure axial-vector current, while curve b gives the maximum when an isoscalar vector current is also present (with gyromagnetic ratio fixed at the quark-model value of -0.12), both plotted versus the effective axial-vector renormalization constant g_A^S. Deviation of g_A^S from the octet value of 0.45 of course requires the presence of a contribution from the SU(3)-singlet axial-vector current. The curves again lie below the BNL data, with a discrepancy exceeding a factor of 2 unless $|g_A^S| \gtrsim 1.5$, which would imply a sizable ninth axial-vector current contribution and a relatively large ninth current renormalization constant as compared with quark-model expectations.[24]

To summarize, if the observed BNL neutral-current pion production rate is to be interpreted in terms of isoscalar V, A currents, then existing elastic and deep-inelastic constraints require that the neutral current contain either a vector current with anomalously large $|g|$ value, or an appreciable coupling to the ninth [SU(3) singlet] axial-vector current.

We conclude by briefly mentioning two other qualitative features of an isoscalar V, A neutral current which may help to distinguish it from alternative phenomenological neutral-current struc-

tures. First, referring to Table VI we note that the V, A interference terms in νp elastic scattering and in weak pion production (for $W \leq 1.4$ GeV) are given by

$$\begin{aligned}
&\text{elastic scattering:} \\
&\quad V,A \text{ interference} \propto (\text{positive}) \\
&\qquad\qquad \times \lambda_3\lambda_4(1+g), \\
&\text{weak pion production:} \\
&\quad V,A \text{ interference} \propto (\text{positive}) \\
&\qquad\qquad \times \lambda_3\lambda_4(0.64+g), \\
&\quad g=\lambda_5/\lambda_4 .
\end{aligned} \tag{65}$$

Hence, except for the small range of isoscalar gyromagnetic ratios

$$-1 \leq g \leq -0.64, \tag{66}$$

the interference terms in neutral-current elastic νp scattering and weak pion production have the same sign. That is, except for g values in the range of Eq. (66), constructive V, A interference in $\nu + N \to \nu + N + \pi$ implies constructive interference in $\nu + p \to \nu + p$ and vice versa. A second useful remark (which has been made by many authors) is that if ν and $\bar{\nu}$ neutral-current cross sections differ (implying the presence of V, A interference effects in a V, A current picture), then the neutral interaction may induce parity-violating terms in the pp, ep, and μp interactions. The significance of this statement is that the same connection between ν, $\bar{\nu}$ cross-section differences and parity-violating effects does not hold in other neutral-current phenomenologies, such as the S, P, T current picture to be discussed in the second paper[5] of this series.

ACKNOWLEDGMENTS

I wish to thank S. F. Tuan for stimulating discussions about the structure of neutral currents, S. B. Treiman for many helpful critical comments in the course of this work, and members of the Argonne-Purdue collaboration and of the Columbia-Illinois-Rockefeller collaboration for conversations about the Argonne and Brookhaven neutrino experiments. Initial parts of this work were done while I was a visitor at the National Accelerator Laboratory; I am grateful to Professor B. W. Lee for the hospitality of the theory group there. I wish also to thank Y. J. Ng for checking the calculation of Appendix B, and H. Frauenfelder, M. L. Goldberger, E. Henley, V. Hughes, J. J. Sakurai, F. J. Wilczek, and L. Wolfenstein for useful remarks.

APPENDIX A

We give here the analog of the low-energy theorem of Eq. (12) for the case of the SU(2) $\otimes$ U(1)-model neutral current of Eq. (52), with $\Delta \mathcal{J}^\lambda = 0$. (The following formulas still neglect the pion mass in the kinematics and so were not used in the numerical work described in the text.) The threshold pion production cross section is given by

$$\begin{aligned}
\frac{1}{|\vec{q}|}\frac{d\sigma(\nu+N\to\nu+N+\pi)}{dt\,dW}\bigg|_{\text{threshold}} &= \frac{G^2}{16\pi^3}\frac{1}{E^2}\left(\frac{g_r}{2M_N{}^2}\right)^2\left(1+\frac{t}{4M_N{}^2}\right)\left(1+\frac{t}{2M_N{}^2}\right)^{-2}\hat{T}, \\
\hat{T} = (2M_N{}^2)^{-1}\Bigg[\left(1+\frac{t}{4M_N{}^2}\right)^{-1}&(H_1{}^2+tH_2{}^2)+(H_4{}^2+tH_3{}^2)\Bigg][4M_N{}^2E^2-t(M_N{}^2+2M_NE)] \\
&+t\left[H_1{}^2+t\left(1+\frac{t}{4M_N{}^2}\right)H_3{}^2\right]+\xi H_1H_3\frac{t}{M_N}(4M_NE-t),
\end{aligned} \tag{A1}$$

with $\xi=1$ (-1) for ν $(\bar{\nu})$-induced reactions, and with

$$\begin{aligned}
H_1 &= a_E^{(-)}(1-2x)2M_N\left(1+\frac{t}{2M_N{}^2}\right)\frac{g_A(k^2)}{g_A}+\frac{t}{2M_N}[\hat{F}_1(k^2)+2M_N\hat{F}_2(k^2)], \\
H_2 &= -\hat{F}_1(k^2)+\frac{t}{2M_N}\hat{F}_2(k^2), \\
H_3 &= a_E^{(-)}\frac{1}{g_A}[F_1^V(k^2)+2M_NF_2^V(k^2)]\frac{1+t/2M_N{}^2}{1+t/4M_N{}^2}+\hat{g}_A(k^2), \\
H_4 &= a_E^{(-)}\frac{2M_N}{g_A}\frac{(1+t/2M_N{}^2)}{(1+t/4M_N{}^2)}\left[F_1^V(k^2)-\frac{t}{2M_N}F_2^V(k^2)\right], \\
\hat{F}_{1,2}(k^2) &= F_{1,2}^V(k^2)[a_E^{(+)}-a_E^{(-)}](1-2x)+F_{1,2}^S(k^2)a_E^{(0)}(-2x), \\
\hat{g}_A(k^2) &= g_A(k^2)[a_E^{(+)}-a_E^{(-)}].
\end{aligned} \tag{A2}$$

TABLE VII. Isospin coefficients appearing in Eq. (A2).

	$a_E^{(+)}$	$a_E^{(-)}$	$a_E^{(0)}$
$\binom{\nu}{\bar\nu}+p\to\binom{\nu}{\bar\nu}+p+\pi^0$	$\frac{1}{2}$	0	$\frac{1}{2}$
$\binom{\nu}{\bar\nu}+n\to\binom{\nu}{\bar\nu}+n+\pi^0$	$\frac{1}{2}$	0	$-\frac{1}{2}$
$\binom{\nu}{\bar\nu}+n\to\binom{\nu}{\bar\nu}+p+\pi^-$	0	$-\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
$\binom{\nu}{\bar\nu}+p\to\binom{\nu}{\bar\nu}+n+\pi^+$	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$

As in the text, we have used the abbreviations $t=-k^2$, $x=\sin^2\theta_W$. The isospin coefficients $a_E^{(\pm 0)}$ are given in Table VII. As the diligent reader may verify, in the case of π^0 production (for which the low-energy-theorem equal-time commutator vanishes) Eqs. (A1) and (A2) reduce to Eq. (12) of the text, with $d\sigma(\nu+N\to\nu+N)/dt$ appropriate to the Weinberg-Salam-model case of Eqs. (21)–(23). Since Eqs. (A1) and (A2) were obtained by algebraic reduction from the Born-approximation expressions of Ref. 6, this agreement provides a cross check on the extensive algebra involved in the extended pion-production model of Sec. III C.

APPENDIX B

We give first a rough estimate of the $O(q)$ corrections in the case of an isoscalar octet axial-vector current. We start from the analogy $\pi^0 \leftrightarrow$ SU(3)-3 index, $\eta \leftrightarrow$ SU(3)-8 index and the fact that $A^{\pi N(+)}=A^{\pi^0 N\to\pi^0 N}$, which allows us to write[48] (in units with $M_\pi=1$)

Isovector:

$$\Delta A_3^{(+)}=\frac{g_A}{g_r}\frac{\partial}{\partial\nu_B}A^{\pi^0 N\to\pi^0 N}\bigg|_0\approx 2.8\,,$$

Octet isoscalar:

$$\Delta A_3^{(\eta)}=\frac{g_A^{(\eta)}}{g_r^{(\eta)}}\frac{\partial}{\partial\nu_B}A^{\pi^0 N\to\eta N}\bigg|_0\,. \tag{B1}$$

Here g_r and $g_r^{(\eta)}$ are, respectively, the $\pi^0 NN$ and the ηNN coupling constants, which according to octet PCAC and SU(3) are related by

$$\frac{g_r^{(\eta)}}{g_r}=\frac{g_A^{(\eta)}}{g_A}\approx\frac{3-(4\times 0.66)}{\sqrt{3}}=0.21\,. \tag{B2}$$

Since the over-all magnitudes of the leading terms in the axial-vector soft-pion amplitude in the isovector and the octet isoscalar cases are governed respectively by g_A and $g_A^{(\eta)}$, a convenient measure of the importance of the $O(q)$ correction in the isoscalar octet axial-vector case, relative to its importance in the isovector axial-vector case, is given by

$$R\equiv\frac{\Delta A_3^{(\eta)}/g_A^{(\eta)}}{\Delta A_3^{(+)}/g_A}=0.16\frac{\partial}{\partial\nu_B}A^{\pi^0 N\to\eta N}\bigg|_0\,. \tag{B3}$$

To estimate the derivative appearing in Eq. (B3), we assume an unsubtracted dispersion relation[47]

$$A^{\pi^0 N\to\eta N}=\frac{1}{\pi}\int_{x_0}^{\infty}dx'\left(\frac{1}{x'-x}+\frac{1}{x'+x+2\nu_B}\right)\times \mathrm{Im}\,A^{\pi^0 N\to\eta N}\,,$$
$$x^0=M_\pi+M_\pi^{\,2}/(2M_N)\,. \tag{B4}$$

We approximate the integral by supposing $A^{\pi^0 N\to\eta N}$ to be dominated by those partial waves containing resonances which couple $\pi^0 N$ strongly to ηN. Referring to the Particle Data Group summary,[46] we see that the only such resonances are the $N^*(1535)$ ($J^P=\frac{1}{2}^-$) and the $N^*(1470)$ ($J^P=\frac{1}{2}^+$). Writing the partial-wave expansion for $A^{\pi^0 N\to\eta N}$ and keeping only the f_{0+} and f_{1-} partial waves to which the $N^*(1535)$ and $N^*(1470)$ respectively contribute, we find[49]

$$\mathrm{Im}\,A^{\pi^0 N\to\eta N}=\frac{4\pi(W+M_N)\,\mathrm{Im}\,f_{0+}^{\pi^0 N\to\eta N}}{[(p_{10}+M_N)(p_{20}+M_N)]^{1/2}}-\frac{4\pi(W-M_N)\,\mathrm{Im}\,f_{1-}^{\pi^0 N\to\eta N}}{[(p_{10}-M_N)(p_{20}-M_N)]^{1/2}}\,, \tag{B5}$$

which is independent of ν_B. So in the approximation of dominance by the $N^*(1535)$ and $N^*(1470)$, we have

$$\frac{\partial}{\partial\nu_B}\mathrm{Im}\,A^{\pi^0 N\to\eta N}=0\,. \tag{B6}$$

Hence, only the derivative of the explicit ν_B in Eq. (B4) contributes, and we find

$$\frac{\partial}{\partial\nu_B}A^{\pi^0 N\to\eta N}\bigg|_0\approx-\frac{2}{\pi}\int_{x_0}^{\infty}\frac{dx'}{(x')^2}\,\mathrm{Im}\,A^{\pi^0 N\to\eta N}(x',\nu_B=0)\,. \tag{B7}$$

Substituting Eq. (B5) into Eq. (B7), we find the bound

$$\left|\frac{\partial}{\partial\nu_B}A^{\pi^0N\to\eta N}\Big|_0\right| \le \frac{2}{\pi}\int_{x_0}^{\infty}\frac{dx'}{(x')^2}\frac{4\pi(W'+M_N)}{[(p'_{10}+M_N)(p'_{20}+M_N)]^{1/2}}|\mathrm{Im}f_{0+}^{\pi^0N\to\eta N}(x')|$$
$$+\frac{2}{\pi}\int_{x_0}^{\infty}\frac{dx'}{(x')^2}\frac{4\pi(W'-M_N)}{[(p'_{10}-M_N)(p'_{20}-M_N)]^{1/2}}|\mathrm{Im}f_{1-}^{\pi^0N\to\eta N}(x')|. \tag{B8}$$

To proceed further, we use resonance dominance to approximate the inelastic amplitude imaginary parts appearing in Eq. (B8) by

$$\mathrm{Im}f_{0+}^{\pi^0N\to\eta N}(x')\approx\frac{g_{0+}^{\eta}}{g_{0+}^{\pi^0}}\mathrm{Im}f_{0+}^{\pi^0N\to\pi^0N}(x'),\quad \mathrm{Im}f_{1-}^{\pi^0N\to\eta N}(x')\approx\frac{g_{1-}^{\eta}}{g_{1-}^{\pi^0}}\mathrm{Im}f_{1-}^{\pi^0N\to\pi^0N}(x'), \tag{B9}$$

with g^{η,π^0} the η, π^0 couplings to the resonance in question. Using the optical theorem to evaluate the right-hand side of Eq. (B9),

$$\mathrm{Im}f^{\pi^0N\to\pi^0N}=\frac{|q_\pi|}{4\pi}\sigma_{\pi^0N}, \tag{B10}$$

combining Eqs. (B8)–(B10) in the narrow-resonance approximation, and expressing the coupling ratios g^η/g^π in terms of resonance partial widths Γ_{η,π^0} and q values $q_{\eta,\pi}$, yields the final formula

$$\left|\frac{\partial}{\partial\nu_B}A^{\pi^0N\to\eta N}\Big|_0\right| \le \frac{2}{\pi}\left[\frac{1}{x^2}\frac{(W+M_N)|q_\pi|}{[(p_{10}+M_N)(p_{20}+M_N)]^{1/2}}\left(\frac{\Gamma_\eta}{\Gamma_{\pi^0}}\frac{|q_\pi|}{|q_\eta|}\right)^{1/2}\int\sigma_{\pi^0N}dx\right]_{N^*(1535)}$$
$$+\frac{2}{\pi}\left[\frac{1}{x^2}\frac{(W-M_N)[(p_{10}+M_N)(p_{20}+M_N)]^{1/2}}{|q_\eta|}\left(\frac{\Gamma_\eta}{\Gamma_{\pi^0}}\frac{|q_\pi|}{|q_\eta|}\right)^{1/2}\int\sigma_{\pi^0N}dx\right]_{N^*(1470)}. \tag{B11}$$

Remembering that the branching ratio of an $I=\frac{1}{2}$ state into π^0 relative to all pionic modes is $\frac{1}{3}$, and taking $|q_\eta|\sim 182$ MeV for the nominally forbidden decay $N^*(1470)-N\eta$ (corresponding to resonance half maximum), we find numerically that the ratio R defined in Eq. (B3) is given by

$$R\lesssim R_{N^*(1535)}+R_{N^*(1470)}=0.014+0.074$$
$$=0.09\,. \tag{B12}$$

Hence the $O(q)$ corrections appear to be substantially less important in the isoscalar octet axial-vector case than they are in the isovector axial-vector case, and so we neglect them. We similarly neglect the $O(q)$ corrections in the unitary singlet axial-vector amplitude, although an analogous argument is not possible in this case since the ninth axial current does not satisfy a PCAC equation.[24] We caution[47] in closing that the above argument is very crude at best, particularly since the $N^*(1470)$ contribution to Eq. (B11) depends as $|q_\eta|^{-3/2}$ on the q_η value assumed for the $N^*(1470)-N\eta$ mode.

We next estimate the extent to which the order-q corrections of Eq. (43a) are already included in the basic pion-production model as a result of unitarization of the (3, 3) multipoles. Using the fact that at $k^2=0$ the electric and longitudinal (3, 3) multipoles are approximately related by[6]

$$\mathcal{L}_{1+}^{(3/2)}\approx-\tfrac{1}{2}\mathcal{E}_{1+}^{(3/2)}, \tag{B13}$$

a simple calculation shows that

$$[A_1^{(-)}-A_1^{(-)B}]|_0\approx-\frac{1}{3}\frac{\mathcal{E}_{1+}^{(3/2)}-\mathcal{E}_{1+}^{(3/2)B}}{|\vec{q}|}\Big|_0,$$
$$[A_3^{(+)}-A_3^{(+)B}]|_0\approx-\frac{4}{3}\frac{\mathcal{E}_{1+}^{(3/2)}-\mathcal{E}_{1+}^{(3/2)B}}{|\vec{q}|}\Big|_0, \tag{B14}$$

where as in the text the superscript B indicates the Born approximation. To evaluate the right-hand side of Eq. (B14), we employ the partial-wave dispersion relation satisfied by $\mathcal{E}_{1+}^{(3/2)}$, which [using Eq. (B13) and making a static nucleon expansion through terms of order M_N^{-1}] takes the simple approximate form

$$\frac{(W+M_N)\mathcal{E}_{1+}^{(3/2)}}{WO_{1+}|\vec{q}|}=\frac{(W+M_N)\mathcal{E}_{1+}^{(3/2)B}}{WO_{1+}|\vec{q}|}+\frac{1}{\pi}\int_{M_\pi}^{\infty}d\omega'\left[\frac{(W'+M_N)\mathrm{Im}\mathcal{E}_{1+}^{(3/2)\prime}}{W'O'_{1+}|\vec{q}'|}\right]\left(\frac{1}{\omega'-\omega}+\frac{11}{36M_N}+\frac{1}{9}\frac{1}{\omega'+\omega}\right), \tag{B15}$$

$$\omega=W-M_N,\quad O_{1+}=[(p_{10}+M_N)(p_{20}+M_N)]^{1/2}.$$

Hence we get

$$\frac{\mathcal{E}_{1+}^{(3/2)}-\mathcal{E}_{1+}^{(3/2)B}}{|\vec{q}|}\Big|_0=\frac{1}{\pi}\int_{M_\pi}^{\infty}d\omega'\left[\frac{M_N}{W'}\frac{(W'+M_N)}{O'_{1+}}\frac{\mathrm{Im}\mathcal{E}_{1+}^{(3/2)\prime}}{|\vec{q}'|}\right]\left(\frac{10}{9}\frac{1}{\omega'}+\frac{11}{36M_N}\right); \tag{B16}$$

integrating the right-hand side numerically, using Eq. (40.22) of Ref. 6 for $\mathcal{S}_{1+}^{(3/2)}$, gives −0.63 in units in which $M_\pi = 1$. Finally, as a point of consistency, we note that a simple calculation shows that Eq. (B16) makes no contribution to the amplitudes $[A_2^{(-)} - A_2^{(-)B}]|_0$ and $[A_4^{(-)} - A_4^{(-)B}]|_0$ which are determined by the zeroth-order PCAC relations of Eq. (42).

APPENDIX C

We give here the nuclear charge-exchange corrections calculation needed to extract the free-nucleon target ratio of Eq. (62) from the measured value of Eqs. (60)-(61). We use the averaged recipe of Eq. (24) of Ref. 40, as extended[50] to the case of nuclei with a neutron excess. In order to simultaneously treat all of the nuclei of current experimental interest,[51] we have performed the calculation outlined in Sec. II B of Ref. 50 using a simple "uniform-well" parameterization of the nuclear density, characterized by a well radius R, a nucleon density ρ, and an rms charge radius a, given by[52]

$$\rho = \frac{A}{\frac{4}{3}\pi R^3},$$
$$R = (\tfrac{5}{3})^{1/2} a, \tag{C1}$$
$$a \approx (0.82 A^{1/3} + 0.58)\ \mathrm{F}.$$

For each nucleus of interest we have calculated two W-averaged charge exchange matrices, one (labeled resonant) appropoiate to the (3, 3) dominated BNL cross section for $\nu + p \to \mu^- + p + \pi^+$, the other (labeled nonresonant) appropriate to the $d\sigma/dW$ curve labeled "Isoscalar pure F_1" in Fig. 7. The results are summarized[53] in Table VIII. In the case of the $I = 0$ nucleus ${}_6\mathrm{C}^{12}$, the resonant matrix of Table VIII implies averaged charge-exchange parameters $\bar{a} = 0.812$, $\bar{d} = 0.137$, $\bar{c} = 0.0392$, in satisfactory agreement with the val-

TABLE VIII. Resonant and nonresonant averaged nuclear charge-exchange matrices for low-invariant-mass ($W \le 1.4$ GeV) weak pion production. The matrix elements are to be read according to the scheme

$$[I] = \begin{bmatrix} I_{++} & I_{+0} & I_{+-} \\ I_{0+} & I_{00} & I_{0-} \\ I_{-+} & I_{-0} & I_{--} \end{bmatrix}.$$

See Appendix C for further details.

T	$[I_T^{\mathrm{res}}]$	$[I_T^{\mathrm{nonres}}]$
${}_6\mathrm{C}^{12}$	$\begin{bmatrix} 0.669 & 0.111 & 0.0318 \\ 0.111 & 0.589 & 0.111 \\ 0.0318 & 0.111 & 0.669 \end{bmatrix}$	$\begin{bmatrix} 0.685 & 0.0866 & 0.0208 \\ 0.0866 & 0.620 & 0.0866 \\ 0.0208 & 0.0866 & 0.685 \end{bmatrix}$
${}_{10}\mathrm{Ne}^{20}$	$\begin{bmatrix} 0.606 & 0.117 & 0.0390 \\ 0.117 & 0.529 & 0.117 \\ 0.0390 & 0.117 & 0.606 \end{bmatrix}$	$\begin{bmatrix} 0.626 & 0.0914 & 0.0257 \\ 0.0914 & 0.560 & 0.0914 \\ 0.0257 & 0.0914 & 0.626 \end{bmatrix}$
${}_9\mathrm{F}^{19}$	$\begin{bmatrix} 0.607 & 0.109 & 0.0348 \\ 0.120 & 0.534 & 0.113 \\ 0.0419 & 0.124 & 0.619 \end{bmatrix}$	$\begin{bmatrix} 0.628 & 0.0854 & 0.0229 \\ 0.0940 & 0.566 & 0.0879 \\ 0.0276 & 0.0968 & 0.637 \end{bmatrix}$
${}_{13}\mathrm{Al}^{27}$	$\begin{bmatrix} 0.565 & 0.113 & 0.0398 \\ 0.121 & 0.494 & 0.116 \\ 0.0453 & 0.124 & 0.574 \end{bmatrix}$	$\begin{bmatrix} 0.588 & 0.0888 & 0.0263 \\ 0.0950 & 0.526 & 0.0908 \\ 0.0300 & 0.0971 & 0.594 \end{bmatrix}$
${}_{35}\mathrm{Br}^{80}$	$\begin{bmatrix} 0.428 & 0.0958 & 0.0397 \\ 0.119 & 0.378 & 0.106 \\ 0.0613 & 0.132 & 0.458 \end{bmatrix}$	$\begin{bmatrix} 0.459 & 0.0777 & 0.0270 \\ 0.0972 & 0.412 & 0.0846 \\ 0.0419 & 0.106 & 0.482 \end{bmatrix}$

ues $\tilde{a}=0.811$, $\tilde{d}=0.138$, $\tilde{c}=0.0450$ given in Table VII of Ref. 40 and obtained by using a "harmonic-well" parameterization of the nuclear density. Given the matrices $[I_T]$ of Table VIII, observed pion-production cross sections are related to free-nucleon cross sections by the following recipe: Let the experimental target contain the mass fractions f_T of the nuclear species with $Z=Z_T$, $A=A_T$. Then the observed cross section per nucleon is given by

$$\begin{bmatrix}\sigma(\text{obs};\pi^+)\\ \sigma(\text{obs};\pi^0)\\ \sigma(\text{obs};\pi^-)\end{bmatrix}=\sum_T f_T[I_T]\begin{bmatrix}\sigma(N_T;\pi^+)\\ \sigma(N_T;\pi^0)\\ \sigma(N_T;\pi^-)\end{bmatrix} \tag{C2}$$

with N_T an effective free-nucleon target given by

$$N_T=\frac{Z_T}{A_T}p+\left(1-\frac{Z_T}{A_T}\right)n. \tag{C3}$$

As an illustration, we apply Eq. (C3) in the BNL case of a mainly carbon and aluminum target. Assuming charged-current pion production to be purely resonant, and neutral-current pion production to be purely nonresonant, we have

$$R_0'=\frac{\sigma(\text{obs};\pi^0\nu)}{2\sigma(\text{obs};\pi^0\mu^-)};$$

$$\sigma(\text{obs};\pi^0\nu)=\sum_{T=\text{C},\text{Al}} f_T\sum_{j=+,0,-}[I_T^{\text{nonres}}]_{0j}\,\sigma(N_T;\pi^j\nu), \tag{C4}$$

$$\sigma(\text{obs};\pi^0\mu^-)=\sum_{T=\text{C},\text{Al}} f_T\sum_{j=+,0,-}[I_T^{\text{res}}]_{0j}\,\sigma(N_T;\pi^j\mu^-).$$

Using the BNL target fractions $f_C=\frac{1}{4}$, $f_{\text{Al}}=\frac{3}{4}$ and assuming an isoscalar neutral current, which implies

$$\sigma(n;\pi^0\nu)=\sigma(p;\pi^0\nu)=\tfrac{1}{2}\sigma(n;\pi^-\nu)=\tfrac{1}{2}\sigma(p;\pi^+\nu), \tag{C5}$$

Eq. (C4) reduces, after some simple algebra to[54]

$$2R_0\equiv\frac{\sigma(n+p;\pi^0\nu)}{\sigma(n;\pi^0\mu^-)}=2R_0'\times0.727(1+0.22r_1+0.23r_2) \tag{C6}$$

with $r_{1,2}$ the charged-current π^+ to π^0 ratios

$$r_1=\frac{\sigma(p;\pi^+\mu^-)}{\sigma(n;\pi^0\mu^-)},\quad r_2=\frac{\sigma(n;\pi^+\mu^-)}{\sigma(n;\pi^0\mu^-)}. \tag{C7}$$

Direct measurements of $r_{1,2}$ in the BNL flux are unavailable, so we have either to use theoretical values for these ratios, or to extrapolate them from the ANL measurements, neglecting possible variations with neutrino energy. The theoretical cross sections tabulated in the fourth and fifth columns of Table III give, respectively,

$$r_1=2.91,\quad r_2=0.88\quad\text{without } O(q)\text{ corrections}, \tag{C8a}$$

$$r_1=4.01,\quad r_2=1.34\quad\text{with } O(q)\text{ corrections}, \tag{C8b}$$

while preliminary ANL data give

$$r_1=3.74\pm0.86,\quad r_2=1.14\pm0.3. \tag{C8c}$$

Substituting into Eq. (C6), Eqs. (C8a)–(C8c) give, respectively,

$$\begin{aligned}2R_0&=2R_0'\times1.59\text{ from (C8a)},\\ 2R_0&=2R_0'\times1.34\text{ from (C8b)},\\ 2R_0&=2R_0'\times1.52\pm0.15\text{ from (C8c)}.\end{aligned} \tag{C9}$$

A charge-exchange correction factor of 1.4 has been assumed in getting Eq. (62) of the text.

---

*Research sponsored in part by the U. S. Atomic Energy Commission under Grant No. AT(11-1)-2220.

[1]F. J. Hasert *et al.*, Phys. Lett. 46B, 138 (1973); A. Benvenuti *et al.*, Phys. Rev. Lett. 32, 800 (1974).

[2]P. A. Schreiner, Argonne National Laboratory Report No. ANL/HEP 7436 (unpublished); S. J. Barish, Bull. Am. Phys. Soc. 20, 86 (1975); D. Carmony (unpublished).

[3]Columbia-Illinois-Rockefeller collaboration, data presented at the Argonne Symposium on Neutral Currents, March 6, 1975 and the Paris Weak Interactions Symposium, March 18-20, 1975.

[4]S. L. Adler, Phys. Rev. Lett. 33, 1511 (1974). In treating the $O(q)$ additions in our original ANL data analysis, we neglected to subtract away the resonant multipole contributions, as we have done following Eq. (43a) of the text. The pion production coefficients of Table V and the discussion of Sec. IV B follow the original analysis, and hence are subject to small corrections (of order 10% in the cross-section bounds). Everywhere else in the paper we use $O(q)$ additions which have the resonant multipole contributions subtracted out, as in Eqs. (43c) and (43d).

[5]S. L. Adler, E. W. Colglazier, Jr., J. B. Healy, I. Karliner, J. Lieberman, Y. J. Ng, and H.-S. Tsao, Phys. Rev. D (to be published); S. L. Adler, R. F. Dashen, J. B. Healy, I. Karliner, J. Lieberman, Y. J. Ng, and H.-S. Tsao, *ibid.* (to be published).

[6]S. L. Adler, Ann. Phys. (N.Y.) 50, 189 (1968). [See also S. L. Adler, Phys. Rev. D 9, 229 (1974).] We have taken the axial-vector form-factor mass as $M_A=0.9$ GeV.

[7]See, for example, S. L. Adler and R. F. Dashen, *Current Algebras* (Benjamin, New York, 1968).

[8]We follow throughout the metric and γ-matrix conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), Appendix A. Also, throughout this paper ν will be understood to mean a muon neutrino ν_μ

[9]For a more detailed discussion, see S. L. Adler and W. I. Weisberger, Phys. Rev. 169, 1392 (1968).

[10]The equal-time commutator term also vanishes for π^0 production by an arbitrary V, A neutral current, and so Eq. (12) holds in this case as well. Vanishing of the equal-time commutator in the isoscalar V, A case was noted by J. J. Sakurai, in *Neutrinos—1974*, proceedings of the Fourth International Conference on Neutrino Physics and Astrophysics, Philadelphia, 1974, edited by C. Baltay (A.I.P., New York, 1974).

[11]The Argonne flux is given, for example, in P. A. Schreiner and F. von Hippel, Argonne National Laboratory Report No. ANL/HEP 7309 (unpublished).

[12]See P. A. Schreiner, Ref. 2.

[13]We make this assumption throughout our discussion of bounds on ANL threshold pion production.

[14]In our discussion of the BNL data in Sec. IV C, the effect of cuts used in setting the relevant CERN bound on $\sigma(\nu+p\to\nu+p)$ will be explicitly taken into account.

[15]B. C. Barish *et al.*, Phys. Rev. Lett. 34, 538 (1975).

[16]We follow here the notation of S. L. Adler, E. W. Colglazier, Jr., J. B. Healy, I. Karliner, J. Lieberman, Y. J. Ng, and H.-S. Tsao, Phys. Rev. D 11, 3309 (1975).

[17]J. D. Bjorken, Phys. Rev. 179, 1547 (1969).

[18]We follow closely a treatment given in unpublished lecture notes of C. G. Callan.

[19]E. A. Paschos and L. Wolfenstein, Phys. Rev. D 7, 91 (1973).

[20]These equations, in the Weinberg-Salam-model context, were first obtained by A. Pais and S. B. Treiman, Phys. Rev. D 6, 2700 (1972).

[21]A succinct review is given in O. Nachtmann, Nucl. Phys. B38, 397 (1972).

[22]See, for example, L. M. Sehgal, Nucl. Phys. B65, 141 (1973).

[23]In equation (38a) we use the fact that for the axial-vector octet, $\alpha = D/(D+F) \approx 0.66$.

[24]In the quark model, one finds $g_A^{(0)}(0) = g_A^{(8)}(0) \approx 0.74$ and $2M_N F_2^{(0)}(0)/F_1^{(0)}(0) = 2M_N F_2^{(8)}(0)/F_1^{(8)}(0) \approx -0.1$. The prediction for $F_2^{(8)}(0)$ is in excellent agreement with experiment [cf. Eq. (38b)], and so it is likely that the quark-model prediction for $F_2^{(0)}(0)$ will also be reliable. Although the quark-model prediction for $g_A^{(8)}(0)$ is in satisfactory accord with experiment [cf. Eq. (38a)] the prediction for $g_A^{(8)}(0)$ may prove unreliable because of the apparently peculiar properties (such as a possible divergence anomaly) of the ninth axial-vector current. For a discussion of issues connected with the ninth axial-current anomaly and further references, see W. A. Bardeen, Nucl. Phys. B75, 246 (1974).

[25]S. L. Adler, Phys. Rev. 137, B1022 (1965); 139, B1638 (1965).

[26]G. F. Chew, M. L. Goldberger, F. E. Low, and Y. Nambu, Phys. Rev. 106, 1345 (1957).

[27]In writing Eqs. (42) and (43) we have followed the notations of Ref. 6, which differ from those of the present paper. Thus, the superscript (0) was used in Ref. 6 to denote matrix elements of the isoscalar electromagnetic current, which would be proportional to amplitudes denoted by (8) in our present notation. Similarly, spacelike k^2 is positive in the notation of Ref. 6, but negative in our present notation. In writing Eqs. (42) we have replaced the off-shell pion-nucleon coupling $g_r(0)$ by the on-shell coupling g_r. In the numerical evaluation of Eq. (43a) we have used $\mu^V \approx 3.70$ and have taken the values of the pion nucleon amplitudes $\overline{B}^{\pi N(-)}$, $(\partial/\partial\nu_B)\overline{A}^{\pi N(+)}$ at the crossing-symmetric point from the tabulation of H. Pilkuhn *et al.*, Nucl. Phys. B65, 460 (1973). The theoretical analysis leading to Eqs. (42) and (43a) is described in detail in Sec. V of Ref. 6. [See particularly Eqs. (5A.21), (5A.22), (5A.9), and (5A.30).]

[28]F. E. Low, Phys. Rev. 110, 974 (1958); S. L. Adler and Y. Dothan, *ibid.* 151, 1267 (1966).

[29]The axial-vector form-factor dipole mass is ≈ 0.90 GeV, while the dipole mass appearing in the vector-current Sachs form factors is 0.84 GeV.

[30]The experimental points are taken from Fig. 10 of B. Musgrave, Argonne National Laboratory Report No. ANL/HEP 7453 (unpublished).

[31]The ANL result for M_A is $M_A = (0.90 \pm 0.10)$ GeV, and we have used the central value of 0.90 GeV in all of the numerical work. Increasing M_A above the central value will bring the theoretical curves in Fig. 1 closer to the experimental points. For example, an M_A of 1.00 GeV gives cross sections 6–9% larger than those in the figure.

[32]The experimental values were obtained from B. Musgrave, private communication.

[33]The histogram was taken from Fig. 2 of P. A. Schreiner and F. von Hippel, Ref. 11.

[34]The histograms were taken from Fig. 16 of B. Musgrave, Ref. 30.

[35]The induced pseudoscalar from factor h_A makes a vanishing contribution to neutral-current reactions.

[36]This has been checked in one case, by comparing formulas obtained from the λ_4, λ_5 terms of Eq. (44) with the corresponding formulas which are obtained when the nucleon isoscalar electromagnetic form factors $F_{1,2}^S(k^2)$ are used in the final two terms of Eq. (44).

[37]S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967); 27, 1688 (1972); A. Salam, in *Elementary Particle Theory: Relativistic Groups and Analyticity* (Nobel Symposium No. 8), edited by N. Svartholm (Almqvist and Wiksell, Stockholm, 1968), p. 367.

[38]The BNL flux table has been furnished to me by W. Y. Lee and L. Litt (private communication).

[39]The error ± 0.06 largely represents systematic uncertainties; the statistical error is considerably smaller (W. Y. Lee, private communication).

[40]S. L. Adler, S. Nussinov, and E. A. Paschos, Phys. Rev. D 9, 2125 (1974).

[41]For example, in the static limit the cross section for (3,3) excitation by $\mathfrak{F}_3^{5\lambda}$ is $0.202/0.263 = 0.77$ times that for (3,3) excitation by $\mathfrak{F}_3^{\lambda}$; see S. L. Adler, Ref. 6; B. W. Lee, Phys. Lett. 40B, 420 (1972).

[42]The uncut elastic cross-section coefficients are not used in the maximization calculation, but are included for completeness. When no cuts are made, $\sigma^{\mathrm{BNL}}(\nu+n\to\mu^-+p) = 0.88\times10^{-38}\ \mathrm{cm}^2$.

[43]D. C. Cundy *et al.*, Phys. Lett. 31B, 478 (1970). The CERN neutrino flux is given in D. H. Perkins, in *Pro-*

ceedings of the Fifth Hawaii Topical Conference in Particle Physics, 1973, edited by P. N. Dobson, Jr., V. Z. Peterson, and S. F. Tuan (Univ. of Hawaii Press, Honolulu, 1974), Fig. 1.6. Note that the absolute magnitude of the flux is irrelevant in flux averaging—only the shape of the spectrum matters.

[44]A preliminary account of this discussion has been given in S. L. Adler, in a talk given at the 1975 Coral Gables Conference, "Orbis Scientia II," 1975 (unpublished), and IAS report. As a result of a programming error, the curves given in Fig. 1 of the Coral Gables talk are too high; the corrected curves appear as Fig. 4 of the present paper.

[45]The curve shown for the pure F_1 case is very similar to the curve obtained for S, P, T coupling mixtures. See Ref. 5 for further details.

[46]Here we are indicating octet isoscalar amplitudes by the superscript (η); note that in Ref. 6 they were denoted by the superscript (0), while in the text of this paper they are denoted by the superscript (8), with (0) used to indicate singlet isoscalar quantities.

[47]Our notation follows that of Sec. III of Ref. 6, with $x=(W^2-M_N^2)/(2M_N)$. Note that writing an unsubtracted dispersion relation is not formally justified by a Regge analysis of $A^{\pi^0 N\to\eta N}$, since the leading Regge trajectory, the A_1 trajectory, has too high an intercept for the integral in Eq. (B4) to converge. However, a similar use of subtracted dispersion relations coupled with resonance dominance arguments gives a correct estimate of the magnitude of the isovector correction of Eqs. (43) and (B1), even though in this case also, the unsubtracted dispersion relation is formally divergent. Hence, our method in the isoscalar case is an heuristic one, motivated by methods which work in the isovector case. I wish to thank M. L. Goldberger for a helpful discussion of the Reggeology of the $\pi^0 N\to\eta N$ amplitude.

[48]See the Appendix of S. L. Adler, Phys. Rev. 137, B1022 (1965).

[50]S. L. Adler, Phys. Rev. D 9, 2144 (1974).

[51]The CERN Gargamelle group has data in freon (CF_3Br), and a bubble-chamber run at BNL in neon has been proposed. I wish to thank P. Musset and C. Baltay for raising the question of extending the calculations of Ref. 50 to other nuclei.

[52]H. R. Collard, L. R. B. Elton, and R. Hofstadter, in *Landolt-Börnstein: Numerical Data and Functional Relationships; Nuclear Radii,* edited by K.-H. Hellwege (Springer, Berlin, 1967), New Series, Group I, Vol. 2.

[53]B. R. Holstein and M. M. Sternheim are currently studying pion production in nuclei induced by incident protons, using the multiple scattering model of Ref. 50 and variants on the model which take nucleon recoil into account in a detailed way. This study should lead to an improved value of the pion absorption cross section σ_{abs}, which when available will be used to recompute the charge-exchange matrices of Table VIII. In calculating Table VIII we have in the interim used the absorption cross section given in Eq. (27) of Ref. 40.

[54]Under our assumption of an isoscalar neutral current, the simple recipe of Eq. (24) of Ref. (40) gives the formula

$$2R=2R_0'\times(1-2\bar{d})\{1+[\bar{d}/(1-2\bar{d})](r_1+r_2)\}.$$

Taking the effective $\bar{d}$ for the BNL target as $\frac{3}{4}d_{AL}+\frac{1}{4}d_C\approx 0.16$, this formula gives

$$2R=2R_0'\times 0.68[1+0.24(r_1+r_2)],$$

a result very similar to Eq. (C6).

PHYSICAL REVIEW D VOLUME 11, NUMBER 11 1 JUNE 1975

Renormalization constants for scalar, pseudoscalar, and tensor currents*

Stephen L. Adler, E. W. Colglazier, Jr., J. B. Healy, Inga Karliner,
Judy Lieberman, Yee Jack Ng, and Hung-Sheng Tsao
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 27 January 1975)

We calculate the renormalization constants describing nucleon and pion matrix elements of scalar, pseudoscalar, and tensor (S, P, T) current densities. For certain of the constants, expressions can be obtained using standard SU_3 and chiral $SU_3 \otimes SU_3$ methods. To get the remaining constants, we employ the quark model with spherically symmetric quark wave functions to relate the S, P, T renormalization constants to known parameters of the usual vector and axial-vector (V, A) currents. We also evaluate the renormalization constants using the MIT "bag" model quark wave functions. We summarize our results in tabular form, compare the results of the various calculational methods used, and attempt to estimate the accuracy of our predictions.

I. INTRODUCTION

A number of recent papers have examined the possibility that neutral currents may involve scalar, pseudoscalar, and tensor (S, P, T) weak couplings in addition to or in place of the usually assumed vector and axial-vector (V, A) Lorentz structures. In particular, expressions have been given for deep inelastic neutrino nucleon scattering[1,2] (using the quark parton model) and for various low-energy nuclear correlations,[1] assuming a completely general Lorentz structure for the weak neutral current. In order to make phenomenological studies of S, P, T weak neutral couplings which simultaneously use deep-inelastic information on the one hand, and exclusive channel or low-energy nuclear results on the other, it is essential to know the renormalization constants describing the nucleon and pion matrix elements of the S, P, T current densities. The purpose of this paper is to estimate these renormalization constants by using various dynamical models of hadron structure. Our results will be applied in a subsequent publication to a detailed analysis, using current-algebra techniques, of soft-pion production by a weak neutral current of arbitrary Lorentz structure.

Within a general quark-model framework, the currents which we study have the form (for S, P, V, A, T structures, respectively)

$$
\begin{aligned}
\mathfrak{F}_j &= \bar{\psi}\,\tfrac{1}{2}\lambda_j\,\psi, \\
\mathfrak{F}_j^5 &= \bar{\psi}\gamma_5\,\tfrac{1}{2}\lambda_j\,\psi, \\
\mathfrak{F}_j^\lambda &= \bar{\psi}\gamma^\lambda\,\tfrac{1}{2}\lambda_j\,\psi, \\
\mathfrak{F}_j^{5\lambda} &= \bar{\psi}\gamma^\lambda\gamma_5\,\tfrac{1}{2}\lambda_j\,\psi, \\
\mathfrak{F}_j^{\lambda\eta} &= \bar{\psi}\sigma^{\lambda\eta}\,\tfrac{1}{2}\lambda_j\,\psi, \\
&\quad j = 0, \ldots, 8
\end{aligned}
\tag{1}
$$

with ψ being the quark field, $\sigma^{\lambda\eta} = (\frac{1}{2}i)[\gamma^\lambda, \gamma^\eta]$, $\lambda_0 = (\frac{2}{3})^{1/2}$, and with $\lambda_{1,\ldots,8}$ being the usual SU_3 matrices. For describing $\Delta S = 0$ neutral current effects, only the $j = 0, 3, 8$ components of the above nonets are relevant. We write the nucleon matrix elements of these components as[3]

$$
\begin{aligned}
\langle N(p_2)|\,\mathfrak{F}_j\,|N(p_1)\rangle &= \mathfrak{N}_N \bar{u}(p_2)\,F_S^{(j)}(k^2)t_j u(p_1), \\
\langle N(p_2)|\,\mathfrak{F}_j^5\,|N(p_1)\rangle &= \mathfrak{N}_N \bar{u}(p_2)F_P^{(j)}(k^2)\gamma_5 t_j u(p_1), \\
\langle N(p_2)|\,\mathfrak{F}_j^\lambda\,|N(p_1)\rangle &= \mathfrak{N}_N \bar{u}(p_2)[F_1^{(j)}(k^2)\gamma^\lambda + iF_2^{(j)}(k^2)\sigma^{\lambda\eta}k_\eta]\,t_j u(p_1), \\
\langle N(p_2)|\,\mathfrak{F}_j^{5\lambda}\,|N(p_1)\rangle &= \mathfrak{N}_N \bar{u}(p_2)[g_A^{(j)}(k^2)\gamma^\lambda\gamma_5 + h_A^{(j)}(k^2)k^\lambda\gamma_5]\,t_j u(p_1), \\
\langle N(p_2)|\,\mathfrak{F}_j^{\lambda\sigma}\,|N(p_1)\rangle &= \mathfrak{N}_N \bar{u}(p_2)\left[T_1^{(j)}(k^2)\sigma^{\lambda\sigma} + \frac{iT_2^{(j)}(k^2)}{M_N}(\gamma^\lambda k^\sigma - \gamma^\sigma k^\lambda) + \frac{iT_3^{(j)}(k^2)}{M_N{}^2}(P^\lambda k^\sigma - P^\sigma k^\lambda)\right] t_j u(p_1) \\
&= \mathfrak{N}_N \bar{u}(p_2)\left[T_1^{(j)}(k^2)\sigma^{\lambda\sigma} + \frac{i\hat{T}_2^{(j)}(k^2)}{M_N}(\gamma^\lambda k^\sigma - \gamma^\sigma k^\lambda) + \frac{T_3^{(j)}(k^2)}{M_N{}^2}(\sigma^{\lambda\nu}k_\nu k^\sigma - \sigma^{\sigma\nu}k_\nu k^\lambda)\right] t_j u(p_1), \\
\hat{T}_2^{(j)}(k^2) &\equiv T_2^{(j)}(k^2) + 2T_3^{(j)}(k^2), \\
k = p_2 - p_1, \quad P &= p_2 + p_1, \quad \mathfrak{N}_N = \left(\frac{M_N}{p_{20}}\,\frac{M_N}{p_{10}}\right)^{1/2}, \\
t_3 = \tfrac{1}{2}\tau_3, \quad t_0 &= \tfrac{1}{2}(\tfrac{2}{3})^{1/2}, \quad t_8 = \tfrac{1}{2}(\tfrac{1}{3})^{1/2}.
\end{aligned}
\tag{2}
$$

In the above expression, τ_3 is the nucleon Pauli isospin matrix and the spinors $\bar{u}(p_2)$, $u(p_1)$ are understood to include nucleon isospinors. The vector and axial-vector form factors defined above are related to the standard nucleon form factors $F_{1,2}^{V,S}(k^2), g_A(k^2), h_A(k^2)$ by

$$F_{1,2}^{(3)}(k^2) = F_{1,2}^{V}(k^2),\quad g_A^{(3)}(k^2) = g_A(k^2),$$
$$F_{1,2}^{(8)}(k^2) = 3F_{1,2}^{S}(k^2),\quad h_A^{(3)}(k^2) = h_A(k^2). \tag{3}$$

The nonvanishing pion matrix elements of the scalar, pseudoscalar, and tensor currents are

$$\langle\pi^a(p_2)|\mathfrak{F}_j|\pi^b(p_1)\rangle = \mathfrak{N}_\pi F_{S\pi}^{(j)}(k^2)\delta^{ab}t_j,\quad j=0,8$$
$$\langle\pi^a(p_2)|\mathfrak{F}_3^{\lambda\sigma}|\pi^b(p_1)\rangle = \mathfrak{N}_\pi \frac{T_\pi^{(3)}(k^2)}{M_N}\epsilon^{ab3}(P^\lambda k^\sigma - P^\sigma k^\lambda), \tag{4}$$
$$\mathfrak{N}_\pi = \frac{1}{(2p_{10}2p_{20})^{1/2}},\quad \epsilon^{123}=1.$$

Our analysis will give values at $k^2=0$ (and, in certain cases, first derivatives at $k^2=0$) for the various form factors which appear in the above expressions. Effectively, the $k^2=0$ values are the strong interaction renormalization constants describing scalar, pseudoscalar, and tensor density couplings to nucleons and pions.

Two principal calculational methods are used in what follows. First, values for certain of the renormalization constants can be obtained by using standard SU_3 and chiral $SU_3 \otimes SU_3$ methods. For the remaining constants, we use the quark model with spherically symmetric quark wave functions to relate the S, P, T renormalization constants (and certain first derivatives at $k^2=0$) to known parameters of the usual V, A currents. We also give a direct calculation in the quark model using the specific quark wave functions obtained in the MIT "bag" model. Our calculational procedures are further briefly described in Sec. II below. Results of the computations are tabulated in Sec. III, while in Sec. IV we compare results obtained by the various calculational methods used and attempt to estimate the accuracy of our predictions.

II. CALCULATIONAL METHODS

A. SU_3 and chiral $SU_3 \otimes SU_3$ predictions

We begin by discussing those renormalization constants which can be determined within the framework of the Gell-Mann–Oakes–Renner (GMOR) model[4] for SU_3 and chiral $SU_3 \otimes SU_3$ breakdown. In this model, the strong interaction Hamiltonian has the form

$$\mathcal{H} = \mathcal{H}_0 + \kappa(\mathfrak{F}_0 + c\mathfrak{F}_8), \tag{5}$$

with $\mathcal{H}_0$ chiral $SU_3 \otimes SU_3$ symmetric and with $\kappa(\mathfrak{F}_0 + c\mathfrak{F}_8)$ a symmetry-breaking term.[4] The parameter κ has the dimension of mass, while the parameter c is determined by the pseudoscalar meson masses to have the value $c \approx -1.25$. Since κ is not fixed in the GMOR model,[5] we can only determine values of the scalar and pseudoscalar renormalization constants relative to any one of them, say, relative to $F_S^{(8)}(0)$.

We begin by getting relations for the scalar density renormalization constants. Within the scalar octet, SU_3 symmetry relates $F_S^{(3)}(0)/F_S^{(8)}(0)$ to $\alpha_{SS} \approx -0.44$, the $D/(D+F)$ value of the baryon octet semistrong mass splitting, giving

$$F_S^{(3)}(0)/F_S^{(8)}(0) = 1/(3-4\alpha_{SS}). \tag{6}$$

The ninth scalar renormalization constant $F_S^{(0)}(0)$ cannot be calculated by SU_3 symmetry, but can be related to the experimentally measurable pion-nucleon "σ term" parameter[8] $\sigma_{\pi NN}$ and the nucleon SU_3 mass splitting parameter Δm defined respectively by

$$\begin{aligned}\sigma_{\pi NN} &= \tfrac{1}{3}(\sqrt{2}+c)\langle N|\sqrt{2}\,\kappa\mathfrak{F}_0 + \kappa\mathfrak{F}_8|N\rangle\\ &\approx 45 \pm 20\ \text{MeV},\\ \Delta m &= \langle N|\kappa\mathfrak{F}_8|N\rangle\\ &= \frac{1}{c}\left[M_N - \tfrac{1}{2}(M_\Lambda + M_\Sigma)\right]\\ &\approx 173\ \text{MeV},\end{aligned} \tag{7}$$

giving

$$\frac{F_S^{(0)}(0)}{F_S^{(8)}(0)} = \frac{1}{2}\left[\frac{3\sigma_{\pi NN}/\Delta m}{\sqrt{2}+c} - 1\right]. \tag{8a}$$

We remark that if $F_S^{(0)}(0)$ and $F_S^{(8)}(0)$ were equal, as is predicted in the quark model, then Eq. (8a) would fix $\sigma_{\pi NN}$ to have the value

$$\begin{aligned}\sigma_{\pi NN} &= \Delta m(\sqrt{2}+c)\\ &\approx 28\ \text{MeV}.\end{aligned} \tag{8b}$$

Finally, we consider the pion scalar coupling $F_{S\pi}^{(8)}(0)$, which can be evaluated relative to $F_S^{(8)}(0)$ by noting that

$$\begin{aligned}\frac{F_{S\pi}^{(8)}(0)}{F_S^{(8)}(0)} &= \frac{2M_\pi\langle\pi|\kappa c\mathfrak{F}_8|\pi\rangle}{\langle N|\kappa c\mathfrak{F}_8|N\rangle}\\ &= \frac{\tfrac{1}{2}(M_\eta^2 + M_\pi^2) - M_\pi^2}{\tfrac{1}{2}(M_\Lambda + M_\Sigma) - M_N}\\ &= \frac{M_\pi^2}{\sqrt{2}+c}\,\frac{1}{\Delta m},\end{aligned} \tag{9}$$

where the final equality is obtained by using Eq. (7) and the GMOR relation $M_\eta^2/M_\pi^2 = (\sqrt{2}-c)/(\sqrt{2}+c)$.

To get relations for the pseudoscalar density re-

normalization constants, we consider axial-vector current divergences in the GMOR model. Taking first the divergence of $\mathfrak{F}_3^{5\lambda}$, we find

$$\partial_\lambda \mathfrak{F}_3^{5\lambda} = -i[F_3^5, \kappa(\mathfrak{F}_0 + c\mathfrak{F}_8)]$$
$$= i\,\frac{\sqrt{2}+c}{\sqrt{3}}\,\kappa\mathfrak{F}_3^5, \tag{10}$$

which when sandwiched between nucleon states implies that

$$2M_N g_A = \frac{\sqrt{2}+c}{\sqrt{3}}\,\kappa F_P^{(3)}(0), \tag{11}$$

with $g_A = g_A^{(3)}(0)$. Rewriting Eq. (7) for Δm as

$$\Delta m = \frac{1}{2}\frac{\kappa}{\sqrt{3}}\,F_S^{(8)}(0) \tag{12}$$

and dividing Eq. (11) by Eq. (12) we get

$$\frac{F_P^{(3)}(0)}{F_S^{(8)}(0)} = \frac{g_A}{\sqrt{2}+c}\,\frac{M_N}{\Delta m}\,. \tag{13}$$

Next we take the divergence of $\mathfrak{F}_8^{5\lambda}$, giving

$$\partial_\lambda \mathfrak{F}_8^{5\lambda} = -i[F_8^5, \kappa(\mathfrak{F}_0 + c\mathfrak{F}_8)]$$
$$= i\kappa\left[\frac{\sqrt{2}-c}{\sqrt{3}}\,\mathfrak{F}_8^5 + \frac{\sqrt{2}\,c}{\sqrt{3}}\,\mathfrak{F}_0^5\right], \tag{14}$$

which when sandwiched between nucleon states gives

$$2M_N g_A^{(8)}(0) = \frac{\sqrt{2}-c}{\sqrt{3}}\,\kappa F_P^{(8)}(0) + \frac{2c}{\sqrt{3}}\,\kappa F_P^{(0)}(0)\,. \tag{15}$$

Using

$$g_A^{(8)}(0) = g_A(3 - 4\alpha_A) \tag{16}$$

[where $\alpha_A = 0.66$ is the $D/(D+F)$ value of the baryon octet axial-vector vertex] and dividing by Eq. (12) gives the second relation

$$(\sqrt{2}-c)F_P^{(8)}(0) + 2cF_P^{(0)}(0) = (3 - 4\alpha_A)\frac{M_N g_A}{\Delta m}\,F_S^{(8)}(0)\,. \tag{17}$$

A second independent relation for $F_P^{(8)}(0)$ and $F_P^{(0)}(0)$ cannot be obtained in the GMOR model. We note, however, that if $F_P^{(8)}(0)$ and $F_P^{(0)}(0)$ were equal, as in the quark model, then Eq. (17) would reduce to

$$F_P^{(8)}(0) = (3 - 4\alpha_A)\,\frac{g_A}{\sqrt{2}+c}\,\frac{M_N}{\Delta m}\,F_S^{(8)}(0)$$
$$= (3 - 4\alpha_A)F_P^{(3)}(0)\,, \tag{18}$$

an analog of the SU_3 relation of Eq. (16). We remark finally that standard pion pole dominance arguments give for the induced pseudoscalar form factor $h_A^{(3)}(k^2)$ the expression

$$h_A^{(3)}(k^2) = \frac{2M_N g_A}{M_\pi^2 - k^2}\,, \tag{19}$$

from which we get

$$h_A^{(3)}(0) = \frac{2M_N g_A}{M_\pi^2}\,. \tag{20}$$

The formulas obtained in this section are listed in column 1 of Tables I and II.

B. Quark model predictions

We next turn to the quark model, within which we can calculate expressions for all of the scalar, pseudoscalar, and tensor renormalization constants, and for certain of the form-factor first derivatives as well. We use for the nucleon the standard spin-internal-symmetry wave functions of the nonrelativistic quark model,[7]

$$|p, s_z = \tfrac{1}{2}\rangle_{QM} = (\tfrac{1}{18})^{1/2}[2|\mathcal{P}\uparrow\mathfrak{N}\downarrow\mathcal{P}\uparrow\rangle + 2|\mathcal{P}\uparrow\mathcal{P}\uparrow\mathfrak{N}\downarrow\rangle + 2|\mathfrak{N}\downarrow\mathcal{P}\uparrow\mathcal{P}\uparrow\rangle - |\mathcal{P}\uparrow\mathcal{P}\downarrow\mathfrak{N}\uparrow\rangle - |\mathcal{P}\uparrow\mathfrak{N}\uparrow\mathcal{P}\downarrow\rangle$$
$$- |\mathcal{P}\downarrow\mathfrak{N}\uparrow\mathcal{P}\uparrow\rangle - |\mathfrak{N}\uparrow\mathcal{P}\downarrow\mathcal{P}\uparrow\rangle - |\mathfrak{N}\uparrow\mathcal{P}\uparrow\mathcal{P}\downarrow\rangle - |\mathcal{P}\downarrow\mathcal{P}\uparrow\mathfrak{N}\uparrow\rangle]\,,\ \text{etc.}, \tag{21}$$

where p denotes the proton and $\mathcal{P}$, $\mathfrak{N}$ denote quarks. In treating the nucleon spatial wave function, we assume three colored quark triplets to be present, with the physical nucleon constructed as a color singlet.[8] The nucleon states are then completely antisymmetric in the color index, and so satisfy Fermi statistics with completely symmetric spatial wave functions, which we form from one-particle quark orbitals. For the quark orbitals in a nucleon we assume a spherically symmetric Dirac wave-function form:

$$\psi(\vec{r}) = \frac{\mathfrak{N}_q}{(4\pi)^{1/2}}\begin{pmatrix} iJ_0(r) \\ -\vec{\sigma}\cdot\hat{r}\,J_1(r)\end{pmatrix}\chi\,, \tag{22}$$

with J_0 and J_1 arbitrary functions of r, with χ being the quark Pauli spinor, and with the normalization constant $\mathfrak{N}_q$ fixed by the condition

$$1 = \int d^3r\,\psi^\dagger(\vec{r})\psi(\vec{r})$$
$$= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}[J_0^2(r) + J_1^2(r)]\,. \tag{23}$$

TABLE I. Nucleon parameters.

Renormalization constant	SU_3 or chiral $SU_3 \otimes SU_3$ prediction	Quark-model prediction, in terms of $I_{1,\ldots,5}$	Quark-model phenomenological relation	Numerical value from column 3 in MIT model	Numerical value from column 4
$F_S^{(8)}(0)$	$\cdots$	$3(1-2I_2)$	$3(-\frac{1}{2}+\frac{9}{10}g_A)$	1.44	1.86
$F_S^{(0)}(0)$	$F_S^{(3)}(0)\frac{1}{2}\left(\frac{3\sigma_{\pi NN}/\Delta m}{\sqrt{2}+c}-1\right)$	$3(1-2I_2)$	$3(-\frac{1}{2}+\frac{9}{10}g_A)$	1.44	1.86
$F_S^{(3)}(0)$	$F_S^{(8)}(0)/(3-4\alpha_{SS})$	$1-2I_2$	$-\frac{1}{2}+\frac{9}{10}g_A$	0.48	0.62
$F_P^{(8)}(0)$	$\left\{(\sqrt{2}-c)F_P^{(8)}(0)+2cF_P^{(0)}(0)\right.$	$\frac{4}{3}M_N I_3$	μ_p	2.64	2.79
$F_P^{(0)}(0)$	$\left.=(3-4\alpha_A)\frac{M_N g_A}{\Delta m}F_S^{(8)}(0)\right\}$	$\frac{4}{3}M_N I_3$	μ_p	2.64	2.79
$F_P^{(3)}(0)$	$F_S^{(8)}(0)\frac{g_A}{\sqrt{2}+c}\frac{M_N}{\Delta m}$	$\frac{20}{9}M_N I_3$	$\frac{5}{3}\mu_p$	4.41	4.65
$F_1^{(8)}(0)$, $F_1^{(0)}(0)$		3	3	3	3
$F_1^{(3)}(0)$		1	1	1	1
$F_2^{(8)}(0)$, $F_2^{(0)}(0)$		$\frac{2}{3}I_3-\frac{3}{2M_N}$	$(\mu_p-3)/(2M_N)$	$-0.36/(2M_N)$	$-0.21/(2M_N)$
$F_2^{(3)}(0)$		$\frac{10}{9}I_3-\frac{1}{2M_N}$	$(\frac{5}{3}\mu_p-1)/(2M_N)$	$3.40/(2M_N)$	$3.65/(2M_N)$
$F_1^{(8)\prime}(0)$, $F_1^{(0)\prime}(0)$		$\frac{1}{2}(I_1+I_5)-\frac{1}{3M_N}I_3+\frac{3}{4M_N^2}$	$\frac{1}{2}r_p^2-\frac{\mu_p}{4M_N^2}+\frac{3}{4M_N^2}$	$\frac{0.25}{M_\pi^2}$	$\frac{0.17}{M_\pi^2}$ [a]
$F_1^{(3)\prime}(0)$		$\frac{1}{6}(I_1+I_5)-\frac{5}{9M_N}I_3+\frac{1}{4M_N^2}$	$\frac{1}{6}r_p^2-\frac{5\mu_p}{12M_N^2}+\frac{1}{4M_N^2}$	$\frac{0.065}{M_\pi^2}$	$\frac{0.035}{M_\pi^2}$
$\mu_p \equiv M_N[F_2^{(3)}(0)+\frac{1}{3}F_2^{(8)}(0)]+1$		$\frac{4}{3}M_N I_3$	μ_p	2.64	Expt: 2.79
$r_p^2 \equiv 3F_1^{(3)\prime}(0)+\frac{3F_2^{(3)}(0)}{2M_N}+F_1^{(8)\prime}(0)+\frac{F_2^{(8)}(0)}{2M_N}$		I_4+I_5	r_p^2	$\frac{0.50}{M_\pi^2}$	Expt: $\frac{0.33}{M_\pi^2}=0.66\ \mathrm{F}^2$

(Continued on following page)

TABLE I (continued)

Renormalization constant	SU_3 or chiral $SU_3 \times SU_3$ prediction	Quark-model prediction, in terms of $I_{1,\ldots,5}$	Quark-model phenomenological relation	Numerical value from column 3 in MIT model	Numerical value from column 4
$g_A^{(8)}(0)$	$(3-4\alpha_A)g_A$	$1-\frac{4}{3}I_2$	$\frac{3}{5}g_A$	0.65	0.74
$g_A^{(0)}(0)$		$1-\frac{4}{3}I_2$	$\frac{3}{5}g_A$	0.65	0.74
$g_A^{(3)}(0)$		$\frac{5}{3}(1-\frac{4}{3}I_2)$	g_A	1.09	Expt: 1.24
$h_A^{(3)}(0)$	$\frac{2M_N}{M_\pi^2}g_A$	$\frac{4}{9}M_N I_5$	$\frac{5}{3}M_N(\frac{1}{6}r_p^2-\frac{3}{5}g_A')$	$\frac{2M_N}{M_\pi^2}\,0.037$	$\frac{2M_N}{M_\pi^2}\,0.017$
$g_A^{(3)\prime}(0)$		$\frac{5}{3}(\frac{1}{6}I_4-\frac{1}{10}I_5)$	g_A'	$\frac{0.066}{M_\pi^2}$	Expt: $\frac{2.48}{(0.90\text{ GeV})^2}=\frac{0.060}{M_\pi^2}$
$g_A^{(8)\prime}(0)/g_A^{(8)}(0)$, $g_A^{(0)\prime}(0)/g_A^{(0)}(0)$			g_A'/g_A		
$T_1^{(8)}(0)$, $T_1^{(0)}(0)$		$1-\frac{2}{3}I_2$	$\frac{1}{2}+\frac{3}{10}g_A$	0.83	0.87
$T_1^{(3)}(0)$		$\frac{5}{3}(1-\frac{2}{3}I_2)$	$\frac{5}{3}(\frac{1}{2}+\frac{3}{10}g_A)$	1.38	1.45
$T_2^{(8)}(0)$, $T_2^{(0)}(0)$		$-\frac{4}{15}M_N^2 I_5$	$M_N^2[-\frac{1}{6}r_p^2+\frac{3}{5}g_A']$	−1.98	−0.88
$T_2^{(3)}(0)$		$-\frac{4}{9}M_N^2 I_5$	$\frac{5}{3}M_N^2[-\frac{1}{6}r_p^2+\frac{3}{5}g_A']$	−3.30	−1.48
$T_3^{(8)}(0)$, $T_3^{(0)}(0)$		$\frac{1}{2}[-2M_N I_3+\frac{4}{15}M_N^2 I_5+\frac{1}{2}-\frac{1}{3}I_2]$	$\frac{1}{2}[-\frac{3}{2}\mu_p-T_2^{(8)}(0)+\frac{1}{2}T_1^{(8)}(0)]$	−0.78	−1.44
$T_3^{(3)}(0)$		$\frac{1}{2}[-\frac{2}{3}M_N I_3+\frac{4}{9}M_N^2 I_5+\frac{5}{6}-\frac{5}{9}I_2]$	$\frac{1}{2}[-\frac{1}{2}\mu_p-T_2^{(3)}(0)+\frac{1}{2}T_1^{(3)}(0)]$	1.34	0.41
$\hat{T}_2^{(8)}(0)$, $\hat{T}_2^{(8)}(0)$		$-2M_N I_3+\frac{1}{2}-\frac{1}{3}I_2$	$-\frac{3}{2}\mu_p+\frac{1}{2}T_1^{(8)}(0)$	−3.54	−3.75
$\hat{T}_2^{(3)}(0)$		$-\frac{2}{3}M_N I_3+\frac{5}{6}-\frac{5}{9}I_2$	$-\frac{1}{2}\mu_p+\frac{1}{2}T_1^{(3)}(0)$	−0.62	−0.66
$T_1^{(3)\prime}(0)$		$\frac{5}{3}(\frac{1}{6}I_4-\frac{1}{30}I_5)$	$\frac{5}{3}[\frac{1}{24}r_p^2+\frac{9}{20}g_A']$	$0.085/M_\pi^2$	$0.068/M_\pi^2$ [b]
$T_1^{(8)\prime}(0)/T_1^{(8)}(0)$, $T_1^{(0)\prime}(0)/T_1^{(0)}(0)$			$T_1^{(3)\prime}(0)/T_1^{(3)}(0)$		

[a] If $F_1^{(8)}(k^2)$ is parameterized in the dipole form $F_1^{(8)}(k^2)/F_1^{(8)}(0)=(1-k^2/M_S^2)^{-2}$, then this value for $F_1^{(8)\prime}(0)$ corresponds to $M_S=0.84$ GeV.
[b] If $T_1^{(3)}(k^2)$ is parameterized in the dipole form $T_1^{(3)}(k^2)/T_1^{(3)}(0)=(1-k^2/M_T^2)^{-2}$, then this value for $T_1^{(3)\prime}(0)$ corresponds to $M_T=0.91$ GeV.

TABLE II. Pion parameters.

Renormalization constant	SU_3 or chiral $SU_3 \otimes SU_3$ prediction	Quark-model prediction, in terms of $I_{1,\ldots,5}$	Quark-model phenomenological relation	Comment
$F_{S\pi}^{(8)}(0)$	$\frac{M_\pi^2}{\sqrt{2}+c}\frac{1}{\Delta m}F_S^{(8)}(0)$	$\frac{4}{3}M_M 3(1-2I_2)$	$\frac{4}{3}M_M F_S^{(8)}(0)$	Equating columns 2 and 4 $\Rightarrow$ $M_M = 0.53$ GeV
$F_{S\pi}^{(0)}(0)$		$\frac{4}{3}M_M 3(1-2I_2)$	$\frac{4}{3}M_M F_S^{(8)}(0)$	
$T_\pi^{(3)}(0)$		$-\frac{2M_N}{3}fI_3$	$-\frac{1}{2}f\mu_p$	For $f=(\frac{2}{3})^{1/4}$, columns 3 and 4 give -1.19 and -1.26, respectively.

The procedure for calculating nucleon renormalization constants is now completely straightforward.[9] We consider the general quark-model current $\mathfrak{F}_\Gamma = \bar\psi\Gamma\psi$ (Γ is a combination of γ and λ matrices) with one-nucleon matrix element

$$\langle N(p_2)|\,\mathfrak{F}_\Gamma\,|N(p_1)\rangle = \mathfrak{N}_N \bar u(p_2)K_\Gamma(p_2,p_1)u(p_1)\,. \quad (24)$$

Working in the brick-wall frame with

$$\vec p_1 = -\tfrac{1}{2}\vec k,\quad \vec p_2 = \tfrac{1}{2}\vec k$$
$$p_{10} = p_{20} = [M_N^2 + \tfrac{1}{4}\vec k^2]^{1/2}, \quad (25)$$

and using our independent-orbital construction of the nucleon wave function, we get the relation

$$\mathfrak{N}_N \bar u(p_2)K_\Gamma(p_2,p_1)u(p_1) = {}_{QM}\Big\langle N\Big|\sum_{\text{quarks}}\mathfrak{M}_\Gamma(\vec k)\Big|N\Big\rangle_{QM}, \quad (26)$$

with $\mathfrak{M}_\Gamma$ a matrix in the quark spin-internal-symmetry space given by

$$\mathfrak{M}_\Gamma(\vec k) = \int d^3r\, e^{i\vec k\cdot\vec r}\frac{\mathfrak{N}_q^2}{4\pi}(-iJ_0(r),\ \vec\sigma\cdot\hat r J_1(r)) \times\Gamma\begin{pmatrix} iJ_0(r) \\ -\vec\sigma\cdot\hat r J_1(r)\end{pmatrix}. \quad (27)$$

Taylor-expanding $e^{i\vec k\cdot\vec r}$ and equating terms of zeroth, first, and second order in $\vec k$ on the left- and right-hand sides of Eq. (26), we get formulas at zero momentum transfer for the form factors appearing in $K_\Gamma(p_2,p_1)$, expressed in terms of integrals over the quark wave function. [In evaluating the order $\vec k^2$ relations we drop nucleon recoil terms of order $\vec k^2/(8M_N^2)$ on the left-hand side of Eq. (26); these terms are relatively small and do not represent a well-defined correction since a description of nucleon recoil has not been built into the quark-model wave functions.] The quark wave-function integrals which appear are linear combinations of the five basic integrals

$$
\begin{aligned}
I_1 &= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}J_0^2(r),\\
I_2 &= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}J_1^2(r),\\
I_3 &= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}rJ_0(r)J_1(r),\\
I_4 &= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}r^2J_0^2(r),\\
I_5 &= \int d^3r\,\frac{\mathfrak{N}_q^2}{4\pi}r^2J_1^2(r).
\end{aligned} \quad (28)
$$

Expressions for the nucleon scalar, pseudoscalar, vector, axial-vector, and tensor renormalization constants and certain form factor derivatives, in terms of $I_1,\ldots,I_5$, are given in column 2 of Table I. Eliminating the integrals $I_{1,\ldots,5}$ in terms of the normalization condition of Eq. (23) and four experimentally measured parameters of the vector and axial-vector currents [we take these as g_A, $g_A' \equiv g_A'(0)$, r_p^2 = proton squared charge radius,[10] $\mu_p/(2M_N)$ = proton magnetic moment] gives the phenomenological relations listed in column 3 of Table I. These relations are valid in any quark model with a spherically symmetric wave function of the form of Eq. (22); for example, they are valid in both the MIT[9] and the SLAC[11] bag models and in the Bogoliubov model,[12] even though these assign the quarks very different looking wave functions.

The procedure for calculating pion renormalization constants is analogous to that used for the nucleon, with a few differences which we briefly describe. Just as for the nucleon, we use for the pion the usual spin-internal-symmetry wave functions of the nonrelativistic quark model,[7]

$$|\pi^+\rangle_{QM} = (\tfrac{1}{2})^{1/2}[|\mathfrak{N}\uparrow\mathcal{P}\downarrow\rangle - |\mathfrak{N}\downarrow\mathcal{P}\uparrow\rangle], \text{ etc.} \tag{29}$$

For the quark wave function we use an analog of Eq. (22),

$$\psi(\vec{r}) = \frac{\mathfrak{N}_q f^{-3/2}}{(4\pi)^{1/2}} \begin{pmatrix} iJ_0(r/f) \\ -\vec{\sigma}\cdot\hat{r}J_1(r/f) \end{pmatrix} \chi, \tag{30}$$

with f being a rescaling factor which reflects the fact that quark orbitals in a pion may have a different radius from those in a nucleon. In the MIT bag model[9] f has the value $(\frac{2}{3})^{1/4} \approx 0.90$, not much different from unity. The antiquark wave function is the same as Eq. (30), with the antiquark contribution to a current with even (odd) charge conjugation equal to $+1$ (-1) times the corresponding quark contribution. The pion analog of Eqs. (24)–(27) is evidently

$$\begin{aligned} \langle \pi(p_2) \| \mathcal{F}_\Gamma | \pi(p_1) \rangle &= \mathfrak{N}_\pi K'_\Gamma(p_2, p_1) \\ &= {}_{QM}\Big\langle \pi \Big| \sum_{\text{quarks}} \mathfrak{M}_\Gamma(f\vec{k}) \Big| \pi \Big\rangle_{QM}, \\ \mathfrak{N}_\pi &= \frac{1}{2(M_\pi^2 + \frac{1}{4}\vec{k}^2)^{1/2}} \end{aligned} \tag{31}$$

with $\mathfrak{M}_\Gamma$ being the same matrix function as defined in Eq. (27). In applying Eq. (31) we only expand out to terms of first order in $\vec{k}$, since neglect of recoil in the case of the pion would be unjustified. To order $\vec{k}$, the normalization factor $\mathfrak{N}_\pi$ is just $1/(2M_\pi)$. In the case of the tensor density coupling to the pion this factor of M_π^{-1} is just cancelled by a corresponding factor of M_π coming from K'_Γ, giving a formula for $T_\pi^{(3)}(0)$ which does not involve the pion mass. On the other hand, in the scalar density case the factor M_π^{-1} survives, giving the relation

$$F_{S\pi}^{(8)}(0) = 4M_\pi(1 - 2I_2), \tag{32}$$

which explicitly involves the pion mass. Since, however, the quark model leads to a degenerate meson 35-plet, instead of having a nearly massless pion, we reinterpret the factor M_π in Eq. (32) as being M_M, a typical quark-model meson mass, and thus write

$$F_{S\pi}^{(8)}(0) = 4M_M(1 - 2I_2). \tag{33}$$

As we will see below in Sec. IV, this interpretation of Eq. (32) is in accord with the chiral $SU_3 \otimes SU_3$ formula for $F_{S\pi}^{(8)}(0)$ obtained above. The results of our analysis in the pion case are given in column 2 of Table II (in terms of the integrals $I_{1,\ldots,5}$) and in column 3 of Table II (in terms of vector and axial-vector current parameters).

We conclude this section by giving expressions for the quark orbitals and the integrals $I_{1,\ldots,5}$ in the MIT bag model,[9] which gives a fairly satisfactory account of the measurable parameters of the vector and axial-vector currents. In this model the quarks in a nucleon are confined to a finite spherical region of space of radius R_0, with orbitals

$$\begin{aligned} &J_0(r) = j_0(\omega r/R_0), \quad J_1(r) = j_1(\omega r/R_0), \quad r \le R_0 \\ &J_0(r) = J_1(r) = 0, \quad r \ge R_0, \\ &\omega = 2.04, \quad R_0 = 0.97 M_\pi^{-1}, \\ &j_0(z) = \frac{\sin z}{z}, \quad j_1(z) = \frac{\sin z}{z^2} - \frac{\cos z}{z}. \end{aligned} \tag{34}$$

Evaluating the integrals $I_{1,\ldots,5}$ we find in the MIT model

$$\begin{aligned} I_1 &= \frac{2\omega - 1}{4(\omega - 1)} = 0.740, \\ I_2 &= \frac{2\omega - 3}{4(\omega - 1)} = 0.260, \\ I_3 &= \frac{R_0}{\omega}\,\frac{4\omega - 3}{8(\omega - 1)} = 0.304R_0, \\ I_4 &= \frac{R_0^2}{24\omega^2(\omega - 1)}(4\omega^3 + 2\omega^2 - 4\omega + 3) \\ &= 0.357R_0^2, \\ I_5 &= \frac{R_0^2}{24\omega^2(\omega - 1)}(4\omega^3 - 10\omega^2 + 20\omega - 15) \\ &= 0.175R_0^2. \end{aligned} \tag{35}$$

III. TABULATION OF RESULTS

In Tables I and II we tabulate our results for the form factors defined in Eq. (2). To recapitulate, the quantities c, Δm, α_{SS}, α_A, and $\sigma_{\pi NN}$, defined above in Sec. II A, have the values

$$\begin{aligned} &c \approx -1.25, \\ &\Delta m \approx 173 \text{ MeV}, \\ &\alpha_{SS} \approx -0.44, \\ &\alpha_A \approx 0.66, \\ &\sigma_{\pi NN} \approx 45 \pm 20 \text{ MeV}, \end{aligned} \tag{36}$$

while the integrals $I_{1,\ldots,5}$ are defined and evaluated in Eqs. (28) and (35). The mass M_M, a typical quark-model meson mass introduced in Eq. (33), is of order 0.6–0.8 GeV while the scale factor f introduced in Eq. (30) is close to unity, with the value $(\frac{2}{3})^{1/4} \approx 0.90$ in the MIT model.[13]

IV. DISCUSSION

We conclude by comparing the results obtained by the various calculational methods described above and by attempting to estimate the reliability of our predictions for the scalar, pseudosca-

lar, and tensor current parameters. We turn our attention first to the isovector pseudoscalar renormalization $F_P^{(3)}(0)$ and the isovector induced pseudoscalar amplitude $h_A^{(3)}(0)$, both of which are pion pole dominated. From chiral $SU_3 \otimes SU_3$ and pion pole dominance we find

$$\frac{F_P^{(3)}(0)}{F_S^{(8)}(0)} = \frac{g_A}{\sqrt{2}+c}\,\frac{M_N}{\Delta m} = 41\,,$$
$$h_A^{(3)}(0) = \frac{2M_N g_A}{M_\pi^2} = \frac{2M_N}{M_\pi^2}\,1.24\,, \tag{37}$$

while the MIT model gives[14]

$$\frac{F_P^{(3)}(0)}{F_S^{(8)}(0)} \approx 3.1\,,\quad h_A^{(3)}(0) = \frac{2M_N}{M_\pi^2}\,0.037\,, \tag{38}$$

both much too small. Evidently, the quark-model predictions for pion pole dominated pseudoscalar quantities behave as if the effective pion mass were

$$\left(\frac{41}{3.1}\right)^{1/2} M_\pi = 0.51 \text{ GeV from } F_P^{(3)}(0)\,,$$
$$\left(\frac{1.24}{0.037}\right)^{1/2} M_\pi = 0.81 \text{ GeV from } h_A^{(3)}(0)\,, \tag{39}$$

not unreasonable values since the quark model does not predict an almost massless pion, but rather gives a pion degenerate with all other pseudoscalar and vector mesons in the 35 representation of SU_6. [In fact, the same MIT model calculation giving the value $f = (\frac{2}{3})^{1/4}$ used in Eq. (30) above leads to a value of the 35 representation central mass of $8\omega/(3fR_0) = 0.87$ GeV, consistent with the above estimates.] Referring to Table II, we see that these values for the effective quark-model pion mass are compatible with the value 0.53 GeV obtained by equating the chiral $SU_3 \otimes SU_3$ with the quark-model predictions for the pion scalar density coupling $F_{S\pi}^{(8)}(0)$.

We consider next the isoscalar pseudoscalar renormalization constants $F_P^{(8)}(0)$ and $F_P^{(0)}(0)$. As we have seen, chiral $SU_3 \otimes SU_3$ gives a single equation [Eq. (17)] relating these two constants to $F_S^{(8)}(0)$, which reduces, when $F_P^{(8)}(0)$ and $F_P^{(0)}(0)$ are equal (as in the quark model), to the simple relation

$$\frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = (3-4\alpha_A)\,\frac{F_P^{(3)}(0)}{F_S^{(8)}(0)} = 15\,. \tag{40}$$

This prediction is evidently in serious disagreement with the quark-model value

$$\frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = 1.5\,. \tag{41}$$

The trouble here is most likely the quark-model prediction that $F_P^{(0)}(0) = F_P^{(8)}(0)$, which leads to near cancellation of the two terms on the left-hand side of Eq. (17) and hence to a large prediction for $F_P^{(8)}(0)$. In actual fact, since there is no light ninth pseudoscalar meson associated with the SU_3-singlet axial-vector current it is likely that $F_P^{(0)}(0) < F_P^{(8)}(0)$. Rewriting Eq. (17) in terms of the ratio

$$r = \frac{F_P^{(0)}(0)}{F_P^{(8)}(0)} \tag{42}$$

we find

$$\frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = \frac{(3-4\alpha_A)M_N g_A}{\Delta m(\sqrt{2}-c+2cr)}\,, \tag{43}$$

which gives the following predictions for $r = 0.3, 0.5, 0.7$ respectively:

$$r = 0.3:\quad \frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = 1.27,\quad \frac{F_P^{(0)}(0)}{F_S^{(8)}(0)} = 0.38\,,$$
$$r = 0.5:\quad \frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = 1.71,\quad \frac{F_P^{(0)}(0)}{F_S^{(8)}(0)} = 0.86\,, \tag{44}$$
$$r = 0.7:\quad \frac{F_P^{(8)}(0)}{F_S^{(8)}(0)} = 2.65,\quad \frac{F_P^{(0)}(0)}{F_S^{(8)}(0)} = 1.85\,,$$

in reasonable agreement with the quark-model value of Eq. (41).

Continuing our comparison of columns 1 and 2 of Table I, we note that SU_3 predicts

$$F_S^{(8)}(0)/F_S^{(3)}(0) = 3 - 4\alpha_{SS} \approx 4.76\,, \tag{45}$$

with an empirical semistrong $D/(D+F)$ ratio $\alpha_{SS} \approx -0.44$, while the quark model gives

$$F_S^{(8)}(0)/F_S^{(3)}(0) = 3\,. \tag{46}$$

Evidently, Eq. (46) represents pure F-type baryon octet semistrong mass splitting, a feature which is a well-known shortcoming of the quark model. For the corresponding axial-vector coupling ratio SU_3 predicts

$$g_A^{(8)}(0)/g_A^{(3)}(0) = 3 - 4\alpha_A \tag{47}$$

with an empirical value $\alpha_A \approx 0.66$, while the quark model gives

$$g_A^{(8)}(0)/g_A^{(3)}(0) = 3/5\,, \tag{48}$$

corresponding to a value of α_A of 0.6. Although the quark-model value of α_A is quite good in this case, the fact that Eq. (47) vanishes for $\alpha_A = 0.75$ makes the predicted $g_A^{(8)}$ in the quark model differ by more than 60% from the value obtained from SU_3 and the empirical α_A. Obviously, in doing phenomenological calculations the predictions of column 1 of Table II should be used (where they are available) in preference to the quark-model values.

For the value of $F_S^{(8)}(0)$ and for all of the tensor density parameters, we must rely solely on quark-model predictions since no information is furnished by SU_3 or chiral $SU_3 \otimes SU_3$ alone. Hence it is essential to have some *a priori* estimate of the reliability of the quark-model predictions.

The following five considerations would appear to be important in forming such an estimate.

1. Consistency of the quark model with SU_3 and chiral $SU_3 \otimes SU_3$ predictions, where available. This question has just been discussed in detail above. In the case of $F_S^{(8)}(0)$, the 60% discrepancy between Eq. (45) and Eq. (46) suggests an estimate of 60–90% for the possible quark model uncertainty.

2. Comparison of the quark-model predictions for the vector and axial-vector parameters with their known experimental values. Referring to Table I, we see that the MIT-model predictions for g_A, g'_A, μ_p, and r_p^2 all agree with experiment[15] to within about 30%, suggesting 30–60% as the general level of reliability for quark-model predictions when other factors (such as pion pole dominance, sensitive cancellations, nucleon recoil corrections, or possible large "glue" contributions) are not involved. In particular, this estimate of the quark-model uncertainty might be expected to apply to the tensor renormalization constant $T_1^{(3)}(0)$.[16]

3. Consistency between the predictions in the final two columns in Table I. Column 5, we recall, gives the predictions of the MIT-model wave functions, while column 6 gives the predictions obtained from the quark-model phenomenological relations of column 4, using as input the empirical values of g_A, g'_A, μ_p, and r_p^2. Sensitive cancellations are unlikely to be involved in cases in which the quark-model predictions are relatively large and relatively unvarying from column 5 to column 6, as for example, for $T_1^{(3)}(0)$. On the other hand, when the quark-model predictions are small or strongly varying from column 5 to column 6, as for $T_1^{(8,0)}(0)$, $\hat{T}_2^{(8)}(0)$, and $T_3^{(8,0,3)}(0)$, they may be considerably less reliable than the 30–60% estimated above.

4. Possible importance of neglected nucleon recoil terms. Whereas $F_S^{(8)}(0)$, $T_1^{(8,0,3)}(0)$, and $\hat{T}_2^{(8,0,3)}(0)$ are true static quantities which are insensitive to our neglect of nucleon recoil, expressions for the renormalization constants $T_3^{(8,0,3)}(0)$ are obtained from the second-order term in $\vec{k}$ in Eq. (27) only when nucleon recoil ambiguities are neglected. This introduces an additional source of uncertainty in the quark-model determination of $T_3^{(8,0,3)}(0)$ relative to the uncertainties present in the quark-model determinations of the other renormalization constants.

5. Possible presence of large "glue" contributions. In the quark model only quark contributions to the various current densities are evaluated, while possible contributions from the "glue" which binds the quarks together are ignored. One peculiar feature of tensor densities is the possibility of induced vector meson couplings of the form

$$gF^{\lambda\eta} = g(\partial^\eta A^\lambda - \partial^\lambda A^\eta)\,, \tag{49}$$

with A being a vector meson field. Such couplings can contribute to the induced tensor renormalization constants $T_2(0)$ and $T_3(0)$, while not affecting the value of $T_1(0)$. If all vector gluons carry a color quantum number, then terms like Eq. (49) will be absent in the color-singlet tensor densities of Eq. (1). In this case, the quark-model predictions for $\hat{T}_2^{(0)}(0)$ and $\hat{T}_2^{(8)}(0)$ should, like that for $T_1^{(3)}(0)$, be relatively reliable. On the other hand, if color-singlet–unitary-singlet gluons are present, then the unitary-singlet tensor current $\mathfrak{F}_0^{\lambda\eta}$ could receive important "gluon" contributions from terms of the form of Eq. (49), introducing a possible large uncertainty into the quark-model prediction for $\hat{T}_2^{(0)}(0)$.

Added note. Applying the method of Eqs. (10)–(15) to the ninth axial-vector current $\mathfrak{F}_0^{5\lambda}$ gives the divergence equation

$$\partial_\lambda \mathfrak{F}_0^{5\lambda} = i\kappa(\tfrac{2}{3})^{1/2}(\mathfrak{F}_0^5 + c\mathfrak{F}_8^5)\,, \tag{50}$$

which when sandwiched between nucleon states gives

$$2M_N g_A^{(0)}(0) = \frac{\kappa}{\sqrt{3}}\,[\sqrt{2}\;F_P^{(0)}(0) + cF_P^{(8)}(0)]\,. \tag{51}$$

Dividing Eq. (51) by Eq. (15) then gives the additional chiral $SU_3 \otimes SU_3$ relation

$$\frac{g_A^{(0)}(0)}{g_A^{(8)}(0)} = \frac{\sqrt{2}\,r + c}{\sqrt{2} - c + 2cr}\,, \tag{52}$$

with $r = F_P^{(0)}(0)/F_P^{(8)}(0)$ being the parameter defined in Eq. (42). For $r = 0.3, 0.5, 0.7$, Eq. (52) gives the respective predictions for $g_A^{(0)}(0)/g_A^{(8)}(0)$,

$$\begin{aligned} r = 0.3&: \quad \frac{g_A^{(0)}(0)}{g_A^{(8)}(0)} = -0.43\,,\\ r = 0.5&: \quad \frac{g_A^{(0)}(0)}{g_A^{(8)}(0)} = -0.38\,,\\ r = 0.7&: \quad \frac{g_A^{(0)}(0)}{g_A^{(8)}(0)} = -0.28\,, \end{aligned} \tag{53}$$

while for the quark-model value $r = 1$, Eq. (52) reduces to the quark-model prediction that $g_A^{(0)}(0)/g_A^{(8)}(0) = 1$. Equations (50)–(53) are valid only when anomalies are not present. When anomalies appear, the above equations apply to the

axial-vector renormalization $g_A^{(0)}(0)$ associated with the "symmetry generating" ninth current, but this is no longer the same as the axial-vector renormalization for the physical ninth axial-vector current. (See W. A. Bardeen, Ref. 17).

ACKNOWLEDGMENTS

We wish to thank R. F. Dashen and R. L. Jaffe for helpful conversations in the course of this work.

---

*Research sponsored in part by the Atomic Energy Commission, Grant No. AT(11-1) 2220.

[1]B. Kayser, G. T. Garvey, E. Fischbach, and S. P. Rosen, Phys. Lett. B52, 385 (1974).

[2]R. L. Kingsley, F. Wilczek, and A. Zee, Phys. Rev. D 10, 2216 (1974).

[3]Our metric and γ matrix conventions follow those of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), Appendix A. In writing the nucleon matrix elements in Eq. (2) we use the fact that the currents defined in Eq. (1) are first-class currents; this eliminates certain otherwise allowed form factor structures. See S. Weinberg, Phys. Rev. 112, 1375 (1958).

[4]M. Gell-Mann, R. J. Oakes, and B. Renner, Phys. Rev. 175, 2195 (1968). Following these authors, we assume the primary $SU_3 \otimes SU_3$ symmetry-breaking term in the Hamiltonian to transform as $(3,\bar{3}) \oplus (\bar{3},3)$, and neglect a possible admixture of terms transforming as $(1,8) \oplus (8,1)$.

[5]In the quark model κ is fixed to have the value $\kappa = (12)^{1/2} \Delta m / F_S^{(8)}(0) = 322$ MeV. [See Eqs. (7) and (12) and the entry for $F_S^{(8)}(0)$ in column 6 of Table I.]

[6]We follow the notation of B. Renner, in *Springer Tracts in Modern Physics*, edited by G. Höhler and E. A. Niekisch (Springer, New York, 1972), Vol. 61, p. 121. The numerical value quoted is recommended in the survey of H. Pilkhuhn *et al.*, Nucl. Phys. B65, 460 (1973), but some determinations give a substantially larger value. For a detailed discussion of determinations of $\sigma_{\pi NN}$, see E. Reya, Rev. Mod. Phys. 46, 545 (1974).

[7]J. J. J. Kokkedee, *The Quark Model* (W. A. Benjamin, N. Y., 1969), Chaps. 5 and 6. The arrows $\uparrow$, $\downarrow$ indicate quark spin direction.

[8]H. Fritzsch and M. Gell-Mann, in *Proceedings of the International Conference on Duality and Symmetry in Hadron Physics*, edited by E. Gotsman (Weizmann Science Press, Jerusalem, 1971). Our results do not depend in any essential way on the use of colored fractionally charged quarks, and would be the same in the Han-Nambu model or any other version of the quark model with a spatially symmetric wave function.

[9]We follow closely the procedure of A. Chodos, R. L. Jaffe, K. Johnson, and C. B. Thorn, Phys. Rev. D 10, 2599 (1974), who give the MIT model predictions for g_A, μ_p, and r_p^2. Their paper is based on the MIT "bag" model of A. Chodos *et al.*, Phys. Rev. D 9, 3471 (1974).

[10]In terms of Sachs form factors, $r_p^2 = 6G'_{Ep}(0)$. The value for r_p^2 given in Table I is based on $6G'_{Ep}(0) \approx 12/(0.71\ \text{GeV}^2) \approx (0.81\ \text{F})^2$.

[11]W. A. Bardeen *et al.*, Phys. Rev. D 11, 1094 (1975).

[12]P. N. Bogoliubov, Ann. Inst. Henri Poincaré 8, 163 (1967).

[13]After these calculations were completed, we learned that $h_A^{(3)}(0)$ and $g_A^{(3)\prime}(0)$ have also been calculated in the MIT model by B. Freedman (private communication from R. L. Jaffe).

[14]The quark-model phenomenological relations give similar predictions.

[15]The discrepancies in the case of g'_A and r_p^2 are both in the same direction, suggesting that the MIT wave functions make the nucleon somewhat too large.

[16]In the theoretical framework of the transformation between current and constituent quarks (abstracted from the free quark model), there is a relation between $g_A(0)$ and $T_1(0)$, since the axial charge and certain components of the tensor charge are generators of SU_{6W} currents. However, the relation appears to be sufficiently arbitrary as to be consistent with the results quoted in Table I.

[17]W. A. Bardeen, Nucl. Phys. B75, 246 (1974).

Reprinted from ANNALS OF PHYSICS
All Rights Reserved by Academic Press, New York and London

Vol. 113, No. 2, August 1978
Printed in Belgium

Trace Anomaly of the Stress-Energy Tensor for Massless Vector Particles Propagating in a General Background Metric*

STEPHEN L. ADLER

The Institute for Advanced Study, Princeton, New Jersey 08540

AND

JUDY LIEBERMAN†

Fermi National Accelerator Laboratory, Batavia, Illinois 60510

Received October 4, 1977

We reanalyze the problem of regularization of the stress-energy tensor for massless vector particles propagating in a general background metric, using covariant point separation techniques applied to the Hadamard elementary solution. We correct an error, pointed out by Wald, in the earlier formulation of Adler, Lieberman, and Ng, and find a stress-energy tensor trace anomaly agreeing with that found by other regularization methods.

1. INTRODUCTION

The problem of regularization of the stress-energy tensor for particles propagating in a general background metric has been treated recently by many authors using differing calculational methods [1]. In an earlier paper by Adler *et. al.* [2], the regularization problem was approached by applying covariant point separation techniques to the Hadamard elementary solutions for the vector and scalar wave equations. The results of Ref. [2] appeared to contradict those obtained by other methods [1], in that no stress-energy tensor trace anomaly was found. Recently, however, Wald [3] has found a mistake in the formal arguments of Sec. 5 and Appendix B of Ref. [2] which accounts for the discrepancy. The mistake consists of the assumption, made in Ref. [2], that the local and boundary-condition-dependent parts $G^{(1L)}$ and $G^{(1B)}$ of the Hadamard elementary solution $G^{(1)}$ are individually symmetric in their arguments x and x', as is their sum $G^{(1)}$. Wald [3] has shown that when this assumption is dropped, a reanalysis of the stress-energy tensor in the scalar particle case gives the standard result for the trace anomaly. The purpose of the present paper is to give the corrected

* Research sponsored by the Energy Research and Development Administration, Grant No. E(11-1)-2220.

† Present address: The Harvard Medical School, Vanderbilt Hall, Avenue Louis Pasteur, Boston, Mass. 02115.

0003-4916/78/1132-0294$05.00/0

stress-energy tensor calculation in the vector particle case; again, when the lack of symmetry of $G^{(1L)}_{\lambda\sigma'}$ and $G^{(1B)}_{\lambda\sigma'}$ is correctly taken into account, the standard result for the trace anomaly is obtained. In order to avoid needless repetition of formulas, this paper has been written in the form of a supplement to Ref. [2], with Eq. (N) of Ref. [2] indicated as Eq. (2.N) .

2. The Calculation

According to Eq. (2.4), the vector particle stress-energy tensor $T_{\alpha\beta}$ is a sum of Maxwell, gauge-breaking, and ghost contributions,

$$T_{\alpha\beta} = T^{\mathrm{M}}_{\alpha\beta} + T^{\mathrm{BR}}_{\alpha\beta} + T^{\mathrm{GH}}_{\alpha\beta}. \tag{1}$$

Since the argument of Eqs. (2.43)–(2.45) showing that

$$\langle T^{\mathrm{BR}}_{\alpha\beta}\rangle + \langle T^{\mathrm{GH}}_{\alpha\beta}\rangle = 0 \tag{2}$$

does not depend on splitting the Hadamard elementary solution $G^{(1)}$ into (L) and (B) parts, it is unaffected by Wald's observation, and so we still have, as in Eqs. (2.5)–(2.6),

$$\begin{aligned} \langle T_{\alpha\beta}(x)\rangle &= \langle T^{\mathrm{M}}_{\alpha\beta}(x)\rangle \\ &= (g_\alpha{}^\mu g_\beta{}^\lambda g^{\nu\sigma} - \tfrac{1}{4} g_{\alpha\beta} g^{\mu\lambda} g^{\nu\sigma})\, P_{\mu\nu\lambda\sigma}\,, \\ P_{\mu\nu\lambda\sigma} &= \lim_{x\to x'} \tfrac{1}{2}\langle F_{\mu\nu}(x)\, F_{\lambda\sigma}(x') + F_{\lambda\sigma}(x')\, F_{\mu\nu}(x)\rangle. \end{aligned} \tag{3}$$

Expressing the expectation in Eq. (3) in terms of the vector particle Hadamard elementary solution $G^{(1)}_{\nu\sigma'}$, and splitting $G^{(1)}_{\nu\sigma'}$ into (L) and (B) parts, we get [as in Eq. (2.48)]

$$\begin{aligned} \langle T_{\alpha\beta}(x)\rangle &= \langle T^{(\mathrm{L})}_{\alpha\beta}(x)\rangle + \langle T^{(\mathrm{B})}_{\alpha\beta}(x)\rangle, \\ \langle T^{(\mathrm{L/B})}_{\alpha\beta}(x)\rangle &= (g_\alpha{}^\mu g_\beta{}^\lambda g^{\nu\sigma} - \tfrac{1}{4} g_{\alpha\beta} g^{\mu\lambda} g^{\nu\sigma})\, P^{(\mathrm{L/B})}_{\mu\nu\lambda\sigma}, \\ P^{(\mathrm{L/B})}_{\mu\nu\lambda\sigma} &= 2G^{(\mathrm{L/B})}_{([\mu\nu][\lambda\sigma])}\,, \\ G^{(\mathrm{L/B})}_{\mu\nu\lambda\sigma} &\equiv [D_\mu D_{\lambda'} G^{(1\mathrm{L/B})}_{\nu\sigma'}(x, x')], \end{aligned} \tag{4}$$

with the notation [ ] in the final line denoting the $x' \to x$ coincidence limit. We assume that the highly singular coincidence limit is regularized in such a manner that $\langle T_{\alpha\beta}(x)\rangle$ is finite and covariantly conserved, which implies that

$$D^\alpha\langle T^{(\mathrm{L})}_{\alpha\beta}(x)\rangle = -D^\alpha\langle T^{(\mathrm{B})}_{\alpha\beta}(x)\rangle. \tag{5}$$

An explicit calculation of the right-hand side of Eq. (5), to be given below, shows that

$$D^\alpha\langle T^{(\mathrm{B})}_{\alpha\beta}(x)\rangle = D^\alpha t^{(\mathrm{L})}_{\alpha\beta}(x), \tag{6}$$

with $t^{(L)}_{\alpha\beta}(x)$ a tensor local in the Riemann curvature and its second covariant derivatives. Thus, from Eq. (5), we have

$$D^\alpha[\langle T^{(L)}_{\alpha\beta}(x)\rangle + t^{(L)}_{\alpha\beta}(x)] = D^\alpha[-\langle T^{(B)}_{\alpha\beta}(x)\rangle + t^{(L)}_{\alpha\beta}(x)] = 0, \tag{7}$$

which implies that $\langle T^{(L)}_{\alpha\beta}(x)\rangle + t^{(L)}_{\alpha\beta}(x)$ is a tensor local in the Riemann tensor and its covariant derivatives, which furthermore is covariantly conserved. Since no parameters with dimension of mass appear in the problem, on dimensional grounds this tensor must have the structure

$$\begin{aligned}
\langle T^{(L)}_{\alpha\beta}(x)\rangle + t^{(L)}_{\alpha\beta}(x) &= c_1 I_{\alpha\beta}(x) + c_2 J_{\alpha\beta}(x),\\
I_{\alpha\beta} &= \frac{1}{(-g)^{1/2}}\frac{\delta}{\delta g^{\alpha\beta}}\int d^4x(-g)^{1/2}R^2\\
&= -2g_{\alpha\beta}R_{.\theta}{}^{\theta} + 2R_{.\alpha\beta} - 2RR_{\alpha\beta} + \frac{1}{2}g_{\alpha\beta}R^2, &(8)\\
J_{\alpha\beta} &= \frac{1}{(-g)^{1/2}}\frac{\delta}{\delta g^{\alpha\beta}}\int d^4x(-g)^{1/2}R^{\rho\tau}R_{\rho\tau}\\
&= -\frac{1}{2}g_{\alpha\beta}R_{.\theta}{}^{\theta} + \frac{1}{2}g_{\alpha\beta}R_{\rho\theta}R^{\rho\theta} + R_{,\alpha\beta} - R_{\alpha\beta,\theta}{}^{\theta} - 2R^{\rho\theta}R_{\rho\alpha\theta\beta}, &(8)
\end{aligned}$$

with c_1 and c_2 arbitrary coefficients. The regularized stress-energy tensor thus becomes

$$\langle T_{\alpha\beta}(x)\rangle = c_1 I_{\alpha\beta}(x) + c_2 J_{\alpha\beta}(x) - t^{(L)}_{\alpha\beta}(x) + \langle T^{(B)}_{\alpha\beta}(x)\rangle, \tag{9}$$

which by Eqs. (7) and (8) is automatically covariantly conserved. To evaluate the trace of $\langle T_{\alpha\beta}(x)\rangle$, we note that from Eq. (4) we have $g^{\alpha\beta}\langle T^{(B)}_{\alpha\beta}(x)\rangle = 0$. Hence we get

$$g^{\alpha\beta}\langle T_{\alpha\beta}(x)\rangle = -2(3c_1 + c_2)\, R_{.\theta}{}^{\theta} - g^{\alpha\beta}t^{(L)}_{\alpha\beta}(x), \tag{10}$$

which is the trace anomaly [and, as is evident from Eqs. (43) and (44) below, is nonvanishing irrespective of the value of $3c_1 + c_2$].

In order to evaluate $t^{(L)}_{\alpha\beta}(x)$ we must carefully calculate $D^\alpha\langle T^{(B)}_{\alpha\beta}(x)\rangle$, keeping in mind the fact that while $G^{(1)}_{\nu\sigma'}(x, x')$ is symmetric under the interchange $\nu, x \leftrightarrow \sigma', x'$, $G^{(1L)}_{\nu\sigma'}$ is not symmetric and hence neither is $G^{(1B)}_{\nu\sigma'}$. Following the notation of [2, Appendix B], we write

$$W_{\nu\sigma'}(x, x') = \tfrac{1}{2}G^{(B)}_{\nu\sigma'}(x, x'), \tag{11}$$

which on substitution into Eq. (4) gives

$$D^\mu\langle T^{(B)}_{\mu\lambda}\rangle = g_\lambda{}^\gamma g^{\beta\delta}D^\alpha P^{(B)}_{\alpha\beta\gamma\delta} - \tfrac{1}{4}g^{\alpha\gamma}g^{\beta\delta}D_\lambda P^{(B)}_{\alpha\beta\gamma\delta}, \tag{12}$$

with

$$P^{(B)}_{\alpha\beta\gamma\delta} = [W_{\beta\delta',\alpha\gamma'} + W_{\alpha\gamma',\beta\delta'} - W_{\alpha\delta',\beta\gamma'} - W_{\beta\gamma',\alpha\delta'}]. \tag{13}$$

In evaluating Eq. (12) we can use the equation of motion at x,

$$W_{\nu\sigma',\mu}{}^{\mu} - R_\nu{}^\theta W_{\theta\sigma'} = 0, \tag{14}$$

but we cannot (as was done in [2, Appendix B]) assume symmetry of $W_{\nu\sigma'}$ or the Lorentz gauge condition of Eq. (2.B3). Beginning with the second term on the right-hand side of Eq. (12), and using Synge's theorem, we get

$$D_\lambda P^{(\mathrm{B})}_{\alpha\beta\gamma\delta} = [W_{\alpha\gamma',\beta\delta'\lambda} - W_{\alpha\delta',\beta\gamma'\lambda} + W_{\alpha\gamma',\beta\delta'\lambda'} - W_{\alpha\delta',\beta\gamma'\lambda'} - (\alpha \leftrightarrow \beta)]$$

$$= \left[W_{\alpha\gamma',\beta\delta'\lambda} - W_{\alpha\delta',\beta\gamma'\lambda} - W_{\gamma\alpha',\delta\beta'\lambda'} + W_{\delta\alpha',\gamma\beta'\lambda'} - \begin{pmatrix}\alpha \leftrightarrow \beta \\ \text{or} \\ \alpha' \leftrightarrow \beta'\end{pmatrix}\right] \qquad \text{(a)}$$

$$+ \left[W_{\gamma\alpha',\delta\beta'\lambda'} - W_{\delta\alpha',\gamma\beta'\lambda'} + W_{\alpha\gamma',\beta\delta'\lambda'} - W_{\alpha\delta',\beta\gamma'\lambda'} - \begin{pmatrix}\alpha \leftrightarrow \beta \\ \text{or} \\ \alpha' \leftrightarrow \beta'\end{pmatrix}\right]. \qquad \text{(b)}$$

$$(15)$$

Substituting Eq. (15) into the second term in Eq. (12), relabeling dummy indices, and combining like terms, we get

$$\begin{aligned} &-\tfrac{1}{4} g^{\alpha\gamma} g^{\beta\delta} D_\lambda P^{(\mathrm{B})}_{\alpha\beta\gamma\delta} \\ &\quad = -\tfrac{1}{2}[(D_\lambda - D_{\lambda'})(W^{\gamma}{}_{\gamma',}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta',}{}^{\delta}{}_{\gamma'})] && \text{(a)} \\ &\quad\quad - \tfrac{1}{2}[(W^{\gamma}{}_{\gamma',}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta',}{}^{\delta}{}_{\gamma'} - W^{\delta}{}_{\gamma',}{}^{\gamma}{}_{\delta'} + W^{\delta}{}_{\delta',}{}^{\gamma}{}_{\gamma'})_{,\lambda'}]. && \text{(b)} \end{aligned} \qquad (16)$$

We next consider the first term on the right-hand side of Eq. (12). Again using Synge's theorem, we have

$$\begin{aligned} D^\alpha P^{(\mathrm{B})}_{\alpha\beta\gamma\delta} = [&W_{\beta\delta',\alpha\gamma'}{}^{\alpha} + W_{\alpha\gamma',\beta\delta'}{}^{\alpha} - W_{\alpha\delta',\beta\gamma'}{}^{\alpha} - W_{\beta\gamma',\alpha\delta'}{}^{\alpha} && \text{(c)} \\ &+ W_{\beta\delta',\alpha\gamma'}{}^{\alpha'} + W_{\alpha\gamma',\beta\delta'}{}^{\alpha'} - W_{\alpha\delta',\beta\gamma'}{}^{\alpha'} - W_{\beta\gamma',\alpha\delta'}{}^{\alpha'}]. && \text{(d)} \end{aligned} \qquad (17)$$

Using the Ricci identity and Eq. (14) to simplify the first line of Eq. (17), and using the cyclic identity, just as is done in [2, Appendix B]; to simplify the second line of Eq. (17), we get, on substitution into the first term of Eq. (12),

$$\begin{aligned} g_\lambda{}^{\gamma} g^{\beta\delta} D^\alpha P^{(\mathrm{B})}_{\alpha\beta\gamma\delta} = &[W_{\alpha\lambda',}{}^{\alpha\delta}{}_{\delta'} - W_{\alpha\delta',}{}^{\alpha\delta}{}_{\lambda'}] && \text{(c)} \\ &+ \tfrac{1}{2}[(W^{\gamma}{}_{\gamma',}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta',}{}^{\delta}{}_{\gamma'} - W^{\delta}{}_{\gamma',}{}^{\gamma}{}_{\delta'} + W^{\delta}{}_{\delta',}{}^{\gamma}{}_{\gamma'})_{,\lambda'}]. && \text{(d)} \end{aligned} \qquad (18)$$

Term (d) of Eq. (18) cancels term (b) of Eq. (16), giving the result

$$\begin{aligned} D^\mu \langle T^{(\mathrm{B})}_{\mu\lambda} \rangle = &-\tfrac{1}{2}[(D_\lambda - D_{\lambda'})(W^{\gamma}{}_{\gamma',}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta',}{}^{\delta}{}_{\gamma'})] && \text{(a)} \\ &+ [W_{\alpha\lambda',}{}^{\alpha\delta}{}_{\delta'} - W_{\alpha\delta',}{}^{\alpha\delta}{}_{\lambda'}]. && \text{(b)} \end{aligned} \qquad (19)$$

If $W_{\nu\sigma'}(x, x')$ were symmetric under the interchange $\nu x \leftrightarrow \sigma' x'$, expression (a) in Eq. (19) would vanish, and if the gauge condition of Eq. (2.B3) $[W_{\nu\sigma',}{}^{\nu} = -W_{,\sigma'}$ with W a scalar] were valid, expression (b) in Eq. (19) would banish, giving the result

$D^{\mu}\langle T^{(B)}_{\mu\lambda}\rangle = 0$ found in [2, Appendix B]. In fact, as we shall now show, expressions (a) and (b) in Eq. (19) are both nonvanishing, providing the origin of the tensor $t^{(L)}_{\alpha\beta}(x)$ appearing in Eq. (6).

We begin with the evaluation of Eq. (19a). Substituting

$$W_{\nu\sigma'} = \frac{1}{(4\pi)^2}\,\Delta^{1/2} w^{(B)}_{\nu\sigma'} = \frac{1}{(4\pi)^2}\,\Delta^{1/2}(w_{\nu\sigma'} - w^{(L)}_{\nu\sigma'}) \tag{20}$$

and noting that (i) $w_{\nu\sigma'}(x, x')$ is symmetric, and so makes no contribution to Eq. (19a), (ii) in the series $w^{(L)}_{\nu\sigma'} = w^{(L)}_{1\nu\sigma'}\sigma(x, x') + w^{(L)}_{2\nu\sigma'}\sigma(x, x')^2 + \cdots$, only the first term contributes to the coincidence limit in Eq. (19a), and (iii) terms with Δ differentiated make no contribution to the coincidence limit in Eq. (19a), we get

$$\begin{aligned}
&-\frac{1}{2}\,[(D_\lambda - D_{\lambda'})(W^{\gamma}{}_{\gamma',}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta',}{}^{\delta}{}_{\gamma'})]\\
&\quad= \frac{1}{2}\frac{1}{(4\pi)^2}\,[(D_\lambda - D_{\lambda'})((\sigma w_1^{(L)\gamma}{}_{\gamma'})_{,}{}^{\delta}{}_{\delta'} - (\sigma w_1^{(L)\gamma}{}_{\delta'})_{,}{}^{\delta}{}_{\gamma'})]\\
&\quad= \frac{1}{2}\frac{1}{(4\pi)^2}\,[5(D_{\lambda'} - D_\lambda)\, w_1^{(L)\gamma}{}_{\gamma'} - 2D_{\gamma'}w_1^{(L)\gamma}{}_{\lambda'} + 2D^{\delta}w_1^{(L)}{}_{\lambda\delta'}].
\end{aligned} \tag{21}$$

The next step is to evaluate the coincidence limits appearing in Eq. (21), following the procedure used by Wald [3] in the scalar case. Writing the recursion relations (2.20) and (2.38) for $v_{1\alpha\sigma'}$ and $w^{(L)}_{1\alpha\sigma'}$ in the form

$$\begin{aligned}
v_1{}^{\alpha}{}_{\sigma'} + \frac{1}{2}\,s\,\frac{Dv_1{}^{\alpha}{}_{\sigma'}}{ds} &= -\frac{1}{4}\,[\Delta^{-1/2}D_\mu D^\mu(\Delta^{1/2}v_0{}^{\alpha}{}_{\sigma'}) - R^{\alpha\gamma}v_{0\gamma\sigma'}],\\
w_1^{(L)\alpha}{}_{\sigma'} + \frac{1}{2}\,s\,\frac{Dw_1^{(L)\alpha}{}_{\sigma'}}{ds} &= -\frac{1}{2}\,v_1{}^{\alpha}{}_{\sigma'} + \frac{1}{4}\,[\Delta^{-1/2}D_\mu D^\mu(\Delta^{1/2}v_0{}^{\alpha}{}_{\sigma'}) - R^{\alpha\gamma}v_{0\gamma\sigma'}],
\end{aligned} \tag{22}$$

with s the arc length along the geodesic joining x' to x, one finds the unique solution regular at $s = 0$

$$w_1^{(L)\alpha}{}_{\sigma'} = -v_1{}^{\alpha}{}_{\sigma'} - \frac{1}{s^2}\int_0^s g^{\alpha}{}_{\bar\alpha}(s, \bar s)\, v_1{}^{\bar\alpha}{}_{\sigma'}(\bar x, x')\,\bar s\, d\bar s. \tag{23}$$

Taking the coincidence limit of Eq. (23) and its first covariant derivative gives

$$\begin{aligned}
[w_1^{(L)\alpha}{}_{\sigma'}] &= \left(-1 - \frac{1}{s^2}\int_0^s \bar s\, d\bar s\right)[v_1{}^{\alpha}{}_{\sigma'}] = -\frac{3}{2}\,[v_1{}^{\alpha}{}_{\sigma'}],\\
[D_\lambda w_1^{(L)\alpha}{}_{\sigma'}] &= \left(-1 - \frac{d}{ds}\frac{1}{s^2}\int_0^s \bar s^2\, d\bar s\right)[D_\lambda v_1{}^{\alpha}{}_{\sigma'}] = -\frac{4}{3}\,[D_\lambda v_1{}^{\alpha}{}_{\sigma'}].
\end{aligned} \tag{24}$$

Thus, making use of Synge's theorem and the fact that $v_1{}^{\gamma}{}_{\gamma'}$ is a symmetric biscalar function of x and x', which implies $[D_\lambda v_1{}^{\gamma}{}_{\gamma'}] = \frac{1}{2}D_\lambda[v_1{}^{\gamma}{}_{\gamma'}]$, we get the evaluations

$$
\begin{aligned}
[D_{\lambda'} w_1^{(L)\gamma}{}_{\gamma'}] &= -\tfrac{5}{6}D_\lambda[v_1{}^{\gamma}{}_{\gamma'}],\\
[D_\lambda w_1^{(L)\gamma}{}_{\gamma'}] &= -\tfrac{2}{3}D_\lambda[v_1{}^{\gamma}{}_{\gamma'}],\\
[D_{\gamma'} w_1^{(L)\gamma}{}_{\lambda'}] &= \tfrac{4}{3}[D_\alpha(v_1{}^{\alpha}{}_{\lambda'})] - \tfrac{3}{2}D_\alpha[v_1{}^{\alpha}{}_{\lambda'}],\\
[D^\delta w_1^{(L)}{}_{\lambda\delta'}] &= \tfrac{4}{3}[D_\alpha(v_1{}^{\alpha}{}_{\lambda'})] - \tfrac{4}{3}D_\alpha[v_1{}^{\alpha}{}_{\lambda'}].
\end{aligned}
\tag{25}
$$

Substituting these into Eq. (21), we get

$$
\begin{aligned}
&-\frac{1}{2}\,[(D_\lambda - D_{\lambda'})(W^{\gamma}{}_{\gamma'}{}^{\delta}{}_{\delta'} - W^{\gamma}{}_{\delta'}{}^{\delta}{}_{\gamma'})]\\
&\quad = \frac{1}{2}\frac{1}{(4\pi)^2}\left\{-\frac{5}{6}\,D_\lambda[v_1{}^{\gamma}{}_{\gamma'}] + \frac{1}{3}\,D_\alpha[v_1{}^{\alpha}{}_{\lambda'}]\right\}.
\end{aligned}
\tag{26}
$$

We trun next to the evaluation of Eq. (19b). We will need, as auxiliary formulas, some consequences of the gauge condition of Eq. (2.26),

$$
G^{(1)}_{\nu\sigma'}{}_{;}{}^{\nu} + G^{(1)}_{0,\sigma'} = 0. \tag{27}
$$

Substituting the Hadamard formulas

$$
\begin{aligned}
G^{(1)}_{\nu\sigma'} &= \frac{2\Delta^{1/2}}{(4\pi)^2}\left(\frac{2g_{\nu\sigma'}}{\sigma} + v_{\nu\sigma'}\ln\sigma + w_{\nu\sigma'}\right),\\
G^{(1)}_0 &= \frac{2\Delta^{1/2}}{(4\pi)^2}\left(\frac{2}{\sigma} + v\ln\sigma + w\right)
\end{aligned}
\tag{28}
$$

into Eq. (27) and equating to zero the coefficient of $\ln\sigma$ gives

$$
(\Delta^{1/2}v_{\nu\sigma'})_{;}{}^{\nu} + (\Delta^{1/2}v)_{,\sigma'} = 0, \tag{29}
$$

while the remainder gives

$$
2\Delta^{1/2}{}_{,\sigma'} + 2(\Delta^{1/2}g_{\nu\sigma'})_{;}{}^{\nu} + \Delta^{1/2}v\sigma_{,\sigma'} + \Delta^{1/2}v_{\nu\sigma'}\sigma_{;}{}^{\nu} + \sigma(\Delta^{1/2}w)_{,\sigma'} + \sigma(\Delta^{1/2}w_{\nu\sigma'})_{;}{}^{\nu} = 0. \tag{30}
$$

Substituting the series expansions

$$
\begin{aligned}
v_{\nu\sigma'} &= \sum_{n=0}^{\infty} v_{n\nu\sigma'}\sigma^n, &\quad w_{\nu\sigma'} &= \sum_{n=0}^{\infty} w_{n\nu\sigma'}\sigma^n,\\
v &= \sum_{n=0}^{\infty} v_n\sigma^n, & w &= \sum_{n=0}^{\infty} w_n\sigma^n
\end{aligned}
\tag{31}
$$

into Eqs. (29) and (30) and equating coefficients order by order in σ, we get the recursion relations

$$2\Delta^{1/2}{}_{.\sigma'} + 2(\Delta^{1/2}g_{\nu\sigma'})_{.}{}^{\nu} + \Delta^{1/2}v_0\sigma_{.\sigma'} + \Delta^{1/2}v_{0\nu\sigma'}\sigma_{.}{}^{\nu} = 0, \tag{32a}$$

$$\Delta^{1/2}v_n\sigma_{.\sigma'} + \Delta^{1/2}v_{n\nu\sigma'}\sigma_{.}{}^{\nu} = -(1/n)\{(\Delta^{1/2}v_{n-1})_{.\sigma'} + (\Delta^{1/2}v_{n-1\nu\sigma'})_{.}{}^{\nu}\}, \qquad n \geqslant 1, \tag{32b}$$

$$\begin{aligned}\Delta^{1/2}w_n\sigma_{.\sigma'} + \Delta^{1/2}w_{n\nu\sigma'}\sigma_{.}{}^{\nu} &= (1/n^2)\{(\Delta^{1/2}v_{n-1})_{,\sigma'} + (\Delta^{1/2}v_{n-1\nu\sigma'})_{.}{}^{\nu}\}\\ &\quad -(1/n)\{(\Delta^{1/2}w_{n-1})_{.\sigma'} + (\Delta^{1/2}w_{n-1\nu\sigma'})_{.}{}^{\nu}\}, \qquad n \geqslant 1.\end{aligned} \tag{32c}$$

In particular, we will need the coincidence limit of the $n = 2$ case of Eqs. (32b) and (32c), which give

$$\begin{aligned}[v_{1\nu\sigma'}{}_{.}{}^{\nu} + v_{1.\sigma'}] &= 0,\\ [w_{1\nu\sigma'}{}_{.}{}^{\nu} + w_{1.\sigma'}] &= 0,\end{aligned} \tag{33}$$

and the $n = 1$ case of Eqs. (32b), (32c), which combined give

$$(\Delta^{1/2}w_0)_{.\sigma'} + (\Delta^{1/2}w_{0\nu\sigma'})_{.}{}^{\nu} = -\Delta^{1/2}w_1\sigma_{.\sigma'} - \Delta^{1/2}w_{1\nu\sigma'}\sigma_{.}{}^{\nu} - \Delta^{1/2}v_1\sigma_{.\sigma'} - \Delta^{1/2}v_{1\nu\sigma'}\sigma_{.}{}^{\nu}. \tag{34}$$

We now proceed as follows. Substituting Eq. (20) into Eq. (19b), and keeping only those terms in the series expansion of Eq. (31) which make a nonvanishing contribution in the coincidence limit, we get

$$\begin{aligned}&[W_{\alpha\lambda'}{}_{,}{}^{\alpha\delta}{}_{\delta'} - W_{\alpha\delta'}{}_{.}{}^{\alpha\delta}{}_{\lambda'}]\\ &= \frac{1}{(4\pi)^2}[(\Delta^{1/2}w^{(B)}_{0\alpha\lambda'} + \Delta^{1/2}w^{(B)}_{1\alpha\lambda'}\sigma)_{.}{}^{\alpha\delta}{}_{\delta'} - (\Delta^{1/2}w^{(B)}_{0\alpha\delta'} + \Delta^{1/2}w^{(B)}_{1\alpha\delta'}\sigma)_{.}{}^{\alpha\delta}{}_{\lambda'}]\\ &= \frac{1}{(4\pi)^2}[(\Delta^{1/2}w_{0\alpha\lambda'})_{.}{}^{\alpha\delta}{}_{\delta'} - (\Delta^{1/2}w_{0\alpha\delta'})_{.}{}^{\alpha\delta}{}_{\lambda'}] \qquad \text{(a)}\\ &\quad + \frac{1}{(4\pi)^2}[(\Delta^{1/2}w^{(B)}_{1\alpha\lambda'}\sigma)_{.}{}^{\alpha\delta}{}_{\delta'} - (\Delta^{1/2}w^{(B)}_{1\alpha\delta'}\sigma)_{.}{}^{\alpha\delta}{}_{\lambda'}]. \qquad \text{(b)}\end{aligned} \tag{35}$$

where in the next to last line we have replaced $w^{(B)}_{0\alpha\lambda'}$ by $w_{0\alpha\lambda'}$, which is justified since $w^{(L)}_{0\alpha\lambda'} = 0$. To evaluate term (a) in Eq. (35), we substitute Eq. (34), which gives

$$\begin{aligned}\text{(a)} &= \frac{1}{(4\pi)^2}[(-(\Delta^{1/2}w_0)_{,\lambda'} - \Delta^{1/2}w_1\sigma_{,\lambda'} - \Delta^{1/2}w_{1\alpha\lambda'}\sigma_{,}{}^{\alpha} - \Delta^{1/2}v_1\sigma_{,\lambda'} - \Delta^{1/2}v_{1\alpha\lambda'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\delta'}\\ &\quad - (-(\Delta^{1/2}w_0)_{,\delta'} - \Delta^{1/2}w_1\sigma_{.\delta'} - \Delta^{1/2}w_{1\alpha\delta'}\sigma_{,}{}^{\alpha} - \Delta^{1/2}v_1\sigma_{,\delta'} - \Delta^{1/2}v_{1\alpha\delta'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\lambda'}]\\ &= \frac{1}{(4\pi)^2}[-(\Delta^{1/2}w_1\sigma_{,\lambda'} + \Delta^{1/2}w_{1\alpha\lambda'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\delta'} + (\Delta^{1/2}w_1\sigma_{,\delta'} + \Delta^{1/2}w_{1\alpha\delta'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\lambda'}]\\ &\quad + \frac{1}{(4\pi)^2}[-(\Delta^{1/2}v_1\sigma_{,\lambda'} + \Delta^{1/2}v_{1\alpha\lambda'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\delta'} + (\Delta^{1/2}v_1\sigma_{,\delta'} + \Delta^{1/2}v_{1\alpha\delta'}\sigma_{,}{}^{\alpha})_{,}{}^{\delta}{}_{\lambda'}]\end{aligned}$$

$$= -\frac{1}{(4\pi)^2}[(\Delta^{1/2}w_{1\alpha\lambda'})_{,}{}^{\alpha'} - (\Delta^{1/2}w_{1\alpha\lambda'})_{,}{}^{\alpha} - (\Delta^{1/2}w_{1\alpha}{}^{\alpha'})_{,\lambda'}$$
$$+ (\Delta^{1/2}w_{1\lambda\beta'})_{,}{}^{\beta} + 3(\Delta^{1/2}w_1)_{,\lambda'}]$$
$$- \frac{1}{(4\pi)^2}[(\Delta^{1/2}v_{1\alpha\lambda'})_{,}{}^{\alpha'} - (\Delta^{1/2}v_{1\alpha\lambda'})_{,}{}^{\alpha} - (\Delta^{1/2}v_{1\alpha}{}^{\alpha'})_{,\lambda'}$$
$$+ (\Delta^{1/2}v_{1\lambda\beta'})_{,}{}^{\beta} + 3(\Delta^{1/2}v_1)_{,\lambda'}]. \tag{36}$$

Substituting Eq. (33), we can eliminate the terms $(\Delta^{1/2}w_1)_{,\lambda'}$ and $(\Delta^{1/2}v_1)_{,\lambda'}$, which gives finally

$$(a) = -\frac{1}{(4\pi)^2}[(\Delta^{1/2}w_{1\alpha\lambda'})_{,}{}^{\alpha'} - 4(\Delta^{1/2}w_{1\alpha\lambda'})_{,}{}^{\alpha} - (\Delta^{1/2}w_{1\alpha}{}^{\alpha'})_{,\lambda'} + (\Delta^{1/2}w_{1\lambda\beta'})_{,}{}^{\beta}]$$
$$- \frac{1}{(4\pi)^2}[(\Delta^{1/2}v_{1\alpha\lambda'})_{,}{}^{\alpha'} - 4(\Delta^{1/2}v_{1\alpha\lambda'})_{,}{}^{\alpha} - (\Delta^{1/2}v_{1\alpha}{}^{\alpha'})_{,\lambda'} + (\Delta^{1/2}v_{1\lambda\beta'})_{,}{}^{\beta}]. \tag{37}$$

Similarly carrying out the differentiations in term (b) of Eq. (35), we get

$$(b) = \frac{1}{(4\pi)^2}[(\Delta^{1/2}w^{(B)}_{1\alpha\lambda'})_{,}{}^{\alpha'} - 4(\Delta^{1/2}w^{(B)}_{1\alpha\lambda'})_{,}{}^{\alpha} - (\Delta^{1/2}w^{(B)}_{1}{}_{\alpha}{}^{\alpha'})_{,\lambda'} + (\Delta^{1/2}w^{(B)}_{1\lambda\beta'})_{,}{}^{\beta}], \tag{38}$$

which when added to the expression in Eq. (37) gives

$$[W_{\alpha\lambda',}{}^{\alpha\delta}{}_{\delta'} - W_{\alpha\delta',}{}^{\alpha\delta}{}_{\lambda'}] = -\frac{1}{(4\pi)^2}[D^{\alpha'}w^{(L)}_{1\alpha\lambda'} - 4D^{\alpha}w^{(L)}_{1\alpha\lambda'} - D_{\lambda'}w^{(L)}_{1}{}_{\alpha}{}^{\alpha'} + D^{\beta}w^{(L)}_{1\lambda\beta'}]$$
$$- \frac{1}{(4\pi)^2}[D^{\alpha'}v_{1\alpha\lambda'} - 4D^{\alpha}v_{1\alpha\lambda'} - D_{\lambda'}v_{1\alpha}{}^{\alpha'} + D^{\beta}v_{1\lambda\beta'}]. \tag{39}$$

In writing Eq. (39) we have used the fact, noted above, that derivatives of Δ do not contribute in the coincidence limit. Equation (39) can be reduced to final form by using the following relations [some of which were already given in Eq. (25) above]:

$$\begin{aligned}
[D^{\alpha'}v_{1\alpha\lambda'}] &= \tfrac{1}{2}D_\lambda[v_1] + D_\alpha[v_1{}^\alpha{}_{\lambda'}],\\
[D^{\alpha}v_{1\alpha\lambda'}] &= -\tfrac{1}{2}D_\lambda[v_1],\\
[D_{\lambda'}v_{1\alpha}{}^{\alpha'}] &= \tfrac{1}{2}D_\lambda[v_{1\alpha}{}^{\alpha'}],\\
[D^{\beta}v_{1\lambda\beta'}] &= \tfrac{1}{2}D_\lambda[v_1] + D_\alpha[v_1{}^\alpha{}_{\lambda'}],\\
[D^{\alpha'}w^{(L)}_{1\alpha\lambda'}] &= -\tfrac{2}{3}D_\lambda[v_1] - \tfrac{1}{2}D_\alpha[v_1{}^\alpha{}_{\lambda'}],\\
[D^{\alpha}w^{(L)}_{1\alpha\lambda'}] &= \tfrac{2}{3}D_\lambda[v_1],\\
[D_{\lambda'}w^{(L)}_{1}{}_{\alpha}{}^{\alpha'}] &= -\tfrac{5}{6}D_\lambda[v_{1\alpha}{}^{\alpha'}],\\
[D^{\beta}w^{(L)}_{1\lambda\beta'}] &= -\tfrac{2}{3}D_\lambda[v_1] - \tfrac{4}{3}D_\alpha[v_1{}^\alpha{}_{\lambda'}],
\end{aligned} \tag{40}$$

which when substituted into Eq. (39) yield

$$[W_{\alpha\lambda'}{}^{\alpha\delta}{}_{,\delta'} - W_{\alpha\delta'}{}^{\alpha\delta}{}_{,\lambda'}] = \frac{1}{(4\pi)^2}\left\{-\frac{1}{3}D_\lambda[v_{1\gamma}{}^{\gamma'}] + \frac{5}{6}D_\alpha[v_1{}^\alpha{}_{\lambda'}] + D_\lambda[v_1]\right\}. \quad (41)$$

Combining Eqs. (6), (19), (26), and (41), determine the tensor $t_{\alpha\beta}^{(L)}$ to be

$$t_{\alpha\beta}^{(L)}(x) = \frac{1}{(4\pi)^2}\left\{-\frac{3}{4}g_{\alpha\beta}[v_1{}^\gamma{}_{\gamma'}] + [v_{1\alpha\beta'}] + g_{\alpha\beta}[v_1]\right\}. \quad (42)$$

Using Eqs. (2.33)–(2.34) and the formulas of [2, Appendix D] to evaluate the coincidence limits appearing in Eq. (42),

$$[v_{1\alpha\beta'}] = \tfrac{1}{2}a_{2\alpha\beta}, \qquad [v_1] = \tfrac{1}{2}a_2^{V=0},$$

$$a_{2\alpha\beta} = \tfrac{1}{2}(\tfrac{1}{6}g_\alpha{}^\theta R - R_\alpha{}^\theta)(\tfrac{1}{6}g_{\theta\beta}R - R_{\theta\beta}) + \tfrac{1}{30}g_{\alpha\beta}R_{,\theta}{}^\theta - \tfrac{1}{180}g_{\alpha\beta}R^{\rho\theta}R_{\rho\theta}$$

$$+ \tfrac{1}{180}g_{\alpha\beta}R^{\rho\tau\kappa\theta}R_{\rho\tau\kappa\theta} - \tfrac{1}{6}R_{\alpha\beta,\theta}{}^\theta - \tfrac{1}{12}R^{\rho\tau\theta}{}_\alpha R_{\rho\tau\theta\beta}, \quad (43)$$

$$a_2^{V=0} = \tfrac{1}{72}R^2 + \tfrac{1}{30}R_{,\theta}{}^\theta + \tfrac{1}{180}R^{\rho\tau\kappa\theta}R_{\rho\tau\kappa\theta} - \tfrac{1}{180}R^{\rho\theta}R_{\rho\theta},$$

we get as our final results for the regularized stress-energy tensor and its trace anomaly,

$$\langle T_{\alpha\beta}(x)\rangle = c_1 I_{\alpha\beta}(x) + c_2 J_{\alpha\beta}(x) + \langle T_{\alpha\beta}^{(B)}(x)\rangle$$

$$- \frac{1}{(4\pi)^2}\frac{1}{2}\left\{-\frac{3}{4}g_{\alpha\beta}a_{2\gamma}{}^\gamma + a_{2\alpha\beta} + g_{\alpha\beta}a_2^{V=0}\right\}, \quad (44)$$

$$g_{\alpha\beta}\langle T_{\alpha\beta}(x)\rangle = -2(3c_1 + c_2)R_{,\theta}{}^\theta + \frac{1}{(4\pi)^2}(a_{2\gamma}{}^\gamma - 2a_2^{V=0}).$$

Apart from the undetermined multiple of $R_{,\theta}{}^\theta$ arising from the undetermined multiples of $I_{\alpha\beta}$ and $J_{\alpha\beta}$ in $\langle T_{\alpha\beta}(x)\rangle$, the trace anomaly given in Eq. (44) agrees with that found by other calculational methods. We note, in conclusion, that in the present formulation of the regularization calculation the "curvature-dependent modified averaging" prescription of Ref. [2] plays no role, the regularized local part $\langle T_{\alpha\beta}^{(L)}(x)\rangle$ having been identified, by general arguments, to have the value given in Eq. (8).

ACKNOWLEDGMENTS

We wish to thank R. Wald for sending us a prepublication version of Ref. [3], and for helpful correspondence and conversations. One of us (J.L.) wishes to thank the CERN Theoretical Division for hospitality while a portion of this work was being done.

REFERENCES

1. In the vector case, see L. S. BROWN AND J. P. CASSIDY, *Phys. Rev. D* **15** (1977), 2810; H.-S. TSAO, *Phys. Letters* **68** *B* (1977), 79 and other references cited therein.
2. S. L. ADLER, J. LIEBERMAN, AND Y. J. NG, *Ann. Phys.* **106** (1977), 279–321.
3. R. M. WALD, On the trace anomaly of a conformally invariant quantum field in curved space-time, to appear.

PHYSICAL REVIEW D VOLUME 18, NUMBER 8 15 OCTOBER 1978

"No-hair" theorems for the Abelian Higgs and Goldstone models

Stephen L. Adler and Robert B. Pearson

The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 24 July 1978)

We examine the question of whether black holes can have associated external massive vector and/or scalar fields, when the masses are produced by spontaneous symmetry breaking. Working throughout in the spherically symmetric case, we show that "no-hair" theorems can be proved for the vector field in the Abelian Higgs model, for an arbitrary $\xi R|\phi|^2$ term in the Higgs Lagrangian, and for the Goldstone scalar field model with $\xi=0$. We also show that a Minkowski-space analog problem does have nontrivial screened charge solutions, indicating that the "no-hair" theorems which we prove are consequences of the stringent conditions at the assumed horizon in the general-relativistic case, not of the interacting field or spontaneous-symmetry-breaking aspects of the problem.

I. INTRODUCTION

One of the striking features of the physics of black holes is the existence of "no-hair" theorems, which state that the only external attributes of a black hole (such as its mass M, angular momentum J, and electric charge Q) are those associated with massless fields admitting conserved flux integrals.[1,2] All other types of fields must decouple, under the assumption of a well-behaved geometry at the horizon. These theorems have been proved for a variety of wave equations, including the massless Dirac field, various massive scalar field theories, and the massive spin-1 Proca field. Our purpose in the present paper is to extend this list of equations studied to include classical wave equations in which masses are generated by spontaneous symmetry breaking. This is particularly important in the vector-meson case, since it is widely believed that if massive spin-1 fields exist, they get their masses through a dynamical mechanism of spontaneous symmetry breaking,[3] rather than kinematically as in the Proca equation. The simplest relevant model is the Abelian Higgs model,[3] and so the main focus of this paper is on the question of whether black holes can have Abelian Higgs "hair." We also give some results for the closely related Goldstone scalar-meson model. For simplicity, we assume spherical symmetry throughout, since we expect that if interesting violations of the "no-hair" theorems were to occur, they would be seen in the spherically symmetric case. We find, in fact, no evidence for such violations, and prove "no-hair" theorems for the cases we study. We believe it likely that our proofs will generalize to the nonspherical case.

II. THE ABELIAN HIGGS MODEL

Before writing down the Abelian Higgs model Lagrangian, we begin with some geometric preliminaries.[4] We assume the general time-independent, spherically symmetric line element

$$ds^2=-e^{2\alpha}dt^2+e^{2\beta}dr^2+r^2(d\theta^2+\sin^2\theta d\phi^2)\,. \tag{1}$$

Using a caret to denote components on the orthonormal basis

$$\begin{aligned}&\tilde\omega^{\hat t}=e^{\alpha}\tilde d t\,,\quad \tilde\omega^{\hat r}=e^{\beta}\tilde d r\,,\\ &\tilde\omega^{\hat\theta}=r\tilde d\theta,\quad \tilde\omega^{\hat\phi}=r\sin\theta\tilde d\phi\,,\end{aligned} \tag{2}$$

and using a prime to indicate differentiation d/dr, the Einstein tensor components for this line element are

$$\begin{aligned}G^{\hat r\hat r}&=\frac{2}{r}e^{-2\beta}\alpha'-r^{-2}(1-e^{-2\beta})\,,\\ G^{\hat t\hat t}&=\frac{2}{r}e^{-2\beta}\beta'+r^{-2}(1-e^{-2\beta})\,,\\ G^{\hat\theta\hat\theta}&=G^{\hat\phi\hat\phi}\\ &=e^{-2\beta}(\alpha''+\alpha'^2-\alpha'\beta'+\alpha'r^{-1}-\beta'r^{-1})\,.\end{aligned} \tag{3}$$

The curvature scalar is

$$\begin{aligned}R=&2r^{-2}(1-e^{-2\beta})+4r^{-1}e^{-2\beta}(\beta'-\alpha')\\ &-2e^{-2\beta}(\alpha''+\alpha'^2-\alpha'\beta')\,,\end{aligned} \tag{4}$$

and the Bianchi identity is

$$(G^{\hat r\hat r})'+\alpha'G^{\hat t\hat t}-\frac{2}{r}G^{\hat\theta\hat\theta}+\left(\alpha'+\frac{2}{r}\right)G^{\hat r\hat r}=0\,. \tag{5}$$

The Abelian Higgs model describes a charged scalar field, with a double-well self-interaction, coupled to an initially massless Abelian gauge field. The Lagrangian density for the model, written in generally covariant form, is

$$\begin{aligned}\mathcal{L}=(-g)^{1/2}[&-\tfrac14F_{\mu\nu}F^{\mu\nu}-d_\mu d^{*\mu}-\xi R|\phi|^2\\ &-h(|\phi|^2-\phi_\infty^2)^2]\,,\end{aligned} \tag{6}$$

with

$$F_{\mu\nu}=\frac{\partial A_\mu}{\partial x^\nu}-\frac{\partial A_\nu}{\partial x^\mu}\,,$$
$$d_\mu=\left(\frac{\partial}{\partial x^\mu}-ieA_\mu\right)\phi\,. \tag{7}$$

The parameter ξ is zero for the usual "minimal" scalar wave equation, while $\xi=\frac{1}{6}$ for the "conformal" scalar wave equation which is conformally invariant in the absence of mass terms. Spontaneous symmetry breaking arises because the effective potential

$$V(\phi)=h(|\phi|^2-\phi_\infty^2)^2 \tag{8}$$

has its minimum at $|\phi|=\phi_\infty$, rather than at $\phi=0$.

Following the analysis of Bekenstein[2] in the similar case of charged scalar electrodynamics, we use the fact that in the time-independent case of interest in black-hole physics we can choose a gauge in which ϕ is real, $A_i=0$, and A_t is time independent.[5] In this gauge the field equations of motion which follow from the Lagrangian of Eq. (6) are

$$(e^{-(\alpha+\beta)}r^2A_t')'=r^2e^{\beta-\alpha}2e^2A_t\phi^2\,, \tag{9a}$$

$$\begin{aligned}&(e^{\alpha-\beta}r^2\phi')'=r^2e^{\alpha+\beta}[-e^{-2\alpha}e^2A_t^2\phi+2h\phi(\phi^2-\phi_\infty^2)]\,,\ \text{"minimal"}\\&(e^{\alpha-\beta}r^2\phi')'=r^2e^{\alpha+\beta}[-e^{-2\alpha}e^2A_t^2\phi+\tfrac{1}{6}R\phi+2h\phi(\phi^2-\phi_\infty^2)]\,,\ \text{"conformal"}\,.\end{aligned} \tag{9b}$$

The stress-energy tensor components in the two cases are the following for the "minimal" model:

$$\begin{aligned}T^{\hat t\hat t}&=\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2+e^{-2\beta}(\phi')^2+e^{-2\alpha}e^2A_t^2\phi^2+h(\phi^2-\phi_\infty^2)^2\,,\\T^{\hat r\hat r}&=-\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2+e^{-2\beta}(\phi')^2+e^{-2\alpha}e^2A_t^2\phi^2-h(\phi^2-\phi_\infty^2)^2\,,\\T^{\hat\theta\hat\theta}&=\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2-e^{-2\beta}(\phi')^2+e^{-2\alpha}e^2A_t^2\phi^2-h(\phi^2-\phi_\infty^2)^2\,,\end{aligned} \tag{10a}$$

and the following for the "conformal" model:

$$\begin{aligned}T&=T^\alpha_\alpha=4h\phi_\infty^2(\phi^2-\phi_\infty^2)\,,\\T^{\hat t\hat t}&=-\tfrac{1}{4}T+\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2+\tfrac{1}{3}e^{-2\beta}(\phi')^2+\tfrac{5}{3}e^{-2\alpha}e^2A_t^2\phi^2\\&\quad+\tfrac{2}{3}\alpha'e^{-2\beta}\phi\phi'-\tfrac{1}{9}\phi^2R+\tfrac{1}{3}\phi^2G^{\hat t\hat t}-\tfrac{1}{3}\phi^2h(\phi^2-\phi_\infty^2)\,,\\T^{\hat r\hat r}&=\tfrac{1}{4}T-\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2+e^{-2\beta}(\phi')^2+\tfrac{1}{3}e^{-2\alpha}e^2A_t^2\phi^2\\&\quad-\tfrac{2}{3}\phi e^{-\beta}(e^{-\beta}\phi')'+\tfrac{1}{9}\phi^2R+\tfrac{1}{3}\phi^2G^{\hat r\hat r}+\tfrac{1}{3}\phi^2h(\phi^2-\phi_\infty^2)\,,\\T^{\hat\theta\hat\theta}&=\tfrac{1}{4}T+\tfrac{1}{2}e^{-2(\alpha+\beta)}(A_t')^2-\tfrac{1}{3}e^{-2\beta}(\phi')^2+\tfrac{1}{3}e^{-2\alpha}e^2A_t^2\phi^2\\&\quad-\tfrac{2}{3}\frac{1}{r}e^{-2\beta}\phi\phi'+\tfrac{1}{9}\phi^2R+\tfrac{1}{3}\phi^2G^{\hat\theta\hat\theta}+\tfrac{1}{3}\phi^2h(\phi^2-\phi_\infty^2)\,.\end{aligned} \tag{10b}$$

In both cases these components satisfy the equation of stress-energy conservation

$$(T^{\hat r\hat r})'+\alpha'T^{\hat t\hat t}-\frac{2}{r}T^{\hat\theta\hat\theta}+\left(\alpha'+\frac{2}{r}\right)T^{\hat r\hat r}=0\,, \tag{11}$$

which determines $T^{\hat\theta\hat\theta}$ given $T^{\hat r\hat r}$ and $T^{\hat t\hat t}$. The two independent Einstein equations are then

$$\begin{aligned}G^{\hat t\hat t}&=8\pi T^{\hat t\hat t}\,,\\G^{\hat r\hat r}&=8\pi T^{\hat r\hat r}\,.\end{aligned} \tag{12}$$

We proceed now to prove a "no-hair" theorem for the Abelian Higgs model. We assume that the coupled system consisting of the vector and the Higgs scalar field and the spherically symmetric space-time geometry, described by Eqs. (1) and (2) above, has a horizon at $r=r_H$ at which all physical scalars are finite. We show that these assumptions imply that the vector field A_t vanishes identically outside the horizon. Multiplying Eq. (9a) by A_t and integrating from r_H to ∞ gives, after an integration by parts,

$$\int_{r_H}^\infty r^2dr[e^{-(\alpha+\beta)}(A_t')^2+2e^{\beta-\alpha}e^2A_t^2\phi^2]=A_tA_t'r^2e^{-(\alpha+\beta)}\Big|_{r_H}^\infty\,. \tag{13}$$

The contribution from ∞ to the right-hand side vanishes, since A_t falls off asymptotically at least as $1/r$. The assumption that the physical scalar $F_{\mu\nu}F^{\mu\nu}$ is bounded at $r=r_H$ implies that $e^{-(\alpha+\beta)}A_t'$ is bounded at the horizon. Hence if $A_t=0$ at r_H, the right-hand side of Eq. (13) vanishes, and the fact that the left-hand side is non-nega-

tive (note that the metric components $e^{2\alpha}$ and $e^{2\beta}$ are non-negative outside the horizon) then implies $A_t \equiv 0$ for all $r \geq r_H$. So we get a "no-hair" theorem unless[6] $A_t|_H \neq 0$.

The remainder of the argument consists of showing that having $A_t|_H \neq 0$ contradicts the assumption that all physical scalars are finite at the horizon.[7] We do this by examining the behavior of the scalar field equation near the horizon. We note first of all that in the "minimal" model boundedness of $T^{\hat{t}\hat{t}}|_H$, and in the "conformal" model boundedness of $T|_H$, both imply that the scalar field ϕ is bounded on the horizon. Hence when $A_t|_H \neq 0$ we have

$$\frac{(0,\frac{1}{6})R\phi + 2h\phi(\phi^2 - \phi_\infty^2)}{-e^{-2\alpha}e^2A_t^2\phi} \sim \frac{e^{2\alpha}}{A_t^2} \times (\text{bounded}) \xrightarrow[r\to r_H]{} 0\ , \tag{14}$$

and the scalar field equation can be approximated near the horizon by

$$(e^{\alpha-\beta}r^2\phi')' + r^2e^{\alpha+\beta}e^{-2\alpha}e^2A_t^2\phi = 0. \tag{15}$$

It proves convenient at this point to change the independent variable from r to λ, with λ the affine parameter of an incoming null geodesic. The differential equation relating λ to r is

$$\begin{aligned} ds^2 = 0 &= -e^{2\alpha}\left(\frac{dt}{d\lambda}\right)^2 + e^{2\beta}\left(\frac{dr}{d\lambda}\right)^2 \\ &= -e^{-2\alpha}P_0^2 + e^{2\beta}\left(\frac{dr}{d\lambda}\right)^2. \end{aligned} \tag{16}$$

Since t is a cyclic variable for a time-independent metric, the conjugate momentum P_0 is a constant of the motion,[8] and so after rescaling λ to make $P_0 = 1$, the second line of Eq. (16) gives

$$\frac{dr}{d\lambda} = e^{-(\alpha+\beta)}. \tag{17}$$

Since the horizon must be a finite affine distance away from any $r > r_H$, the value λ_H of λ at the horizon is finite. In terms of λ, and making the definitions

$$\begin{aligned} q &= e^{2\alpha}\ , \\ p &= e^{-2(\alpha+\beta)}\ , \end{aligned} \tag{18}$$

so that $dr/d\lambda = p^{1/2}$, the approximated scalar field equation becomes

$$\frac{d}{d\lambda}\left(qr^2\frac{d\phi}{d\lambda}\right) + r^2e^2A_t^2q^{-1}\phi = 0\ . \tag{19}$$

To proceed, we need some information on the behavior of q and its derivatives near the horizon. This can be obtained by rearranging Eq. (3) for the Einstein tensor components into the form

$$G^{\hat{r}\hat{r}} + \frac{1}{r^2} = \frac{p^{1/2}}{r}\frac{dq}{d\lambda} + \frac{pq}{r^2}\ , \tag{20a}$$

$$G^{\hat{t}\hat{t}} + G^{\hat{r}\hat{r}} = -\frac{2}{r}q\frac{d}{d\lambda}(p^{1/2})\ , \tag{20b}$$

$$G^{\hat{\theta}\hat{\theta}} = \frac{1}{r}\frac{d}{d\lambda}(p^{1/2}q) + \frac{1}{2}\frac{d^2q}{d\lambda^2}\ . \tag{20c}$$

From the boundedness at the horizon of the left-hand sides of these equations, and the fact that both terms on the right-hand side of Eq. (20a) are non-negative, we deduce the following:

$$\begin{aligned} &\text{Eqs. (20a), (20b)} \Rightarrow pq|_H,\ p^{1/2}\frac{dq}{d\lambda}\Big|_H,\ q\frac{d}{d\lambda}p^{1/2}\Big|_H = \text{bounded} \Rightarrow \frac{d}{d\lambda}(p^{1/2}q)\Big|_H = \text{bounded} \Rightarrow p^{1/2}q|_H = \text{bounded}\ , \\ &\text{Eq. (20c)} \Rightarrow \frac{d^2q}{d\lambda^2}\Big|_H = \text{bounded} \Rightarrow \frac{dq}{d\lambda}\Big|_H = \text{bounded}. \end{aligned} \tag{21}$$

Hence writing

$$\theta = qr^2\ , \tag{22}$$

we have

$$\frac{d\theta}{d\lambda}\Big|_H = r_H^2\frac{dq}{d\lambda}\Big|_H + 2r_Hqp^{1/2}|_H = \text{bounded and} \geq 0,$$

$$\frac{d^2\theta}{d\lambda^2}\Big|_H = r_H^2\frac{d^2q}{d\lambda^2}\Big|_H + 4r_Hp^{1/2}\frac{dq}{d\lambda}\Big|_H + 2r_Hq\frac{dp^{1/2}}{d\lambda}\Big|_H + 2pq|_H = \text{bounded}\ , \tag{23}$$

and in terms of θ the scalar field equation near the horizon takes the compact form

$$\begin{aligned} &\theta\frac{d^2\phi}{d\lambda^2} + \frac{d\theta}{d\lambda}\frac{d\phi}{d\lambda} + \frac{K\phi}{\theta} = 0\ , \\ &K = e^2r^4A_t^2|_H > 0\ . \end{aligned} \tag{24}$$

The final ingredient needed for the argument is the fact that boundedness of $T^{\hat{t}\hat{t}}|_H$ requires

$$q^{-1}\phi^2|_H = \text{bounded} \tag{25a}$$

in the "minimal" model (since in this model all terms in $T^{\hat{t}\hat{t}}$ are non-negative), and

$$\left[q^{-1}\phi^2 + K_1\left(\frac{d\phi}{d\lambda}\right)^2 + K_2\phi\frac{d\phi}{d\lambda}\right]_H = \text{bounded} \tag{25b}$$

in the "conformal" model, with

$$K_1 = [q/(5e^2A_t^2)]_H\,, \quad K_2 = [(dq/d\lambda)/(5e^2A_t^2)]_H \tag{25c}$$

two bounded constants. The strategy of the argument now is to show that Eqs. (23)–(25) are inconsistent. We consider separately the two cases where $d\theta/d\lambda\,|_H > 0$ and where $d\theta/d\lambda\,|_H = 0$.

When $d\theta/d\lambda\,|_H = C > 0$, we can approximate $\theta = C(\lambda - \lambda_H)$ near the horizon, and Eq. (24) takes the form

$$\frac{d^2\phi}{d\lambda^2} + \frac{1}{\lambda-\lambda_H}\frac{d\phi}{d\lambda} + \frac{K}{C^2}\frac{\phi}{(\lambda-\lambda_H)^2} = 0\,, \tag{26}$$

which has the general solution

$$\phi = \phi_0\cos(x+\delta)\,, \quad x = \frac{K^{1/2}}{C}\ln(\lambda-\lambda_H)\,. \tag{27}$$

Hence in this case we find near the horizon

$$\begin{aligned} q^{-1}\phi^2 &+ K_1\left(\frac{d\phi}{d\lambda}\right)^2 + K_2\phi\frac{d\phi}{d\lambda} \\ &\propto (\lambda-\lambda_H)^{-1}(\cos^2 x + C_1\sin^2 x \\ &\qquad + C_2 \sin x\cos x)\,, \end{aligned} \tag{28}$$

which is unbounded at λ_H for all values of the constants $C_{1,2}$. In the second case, when $d\theta/d\lambda\,|_H = 0$, we make an exponential substitution $\phi = e^f$ in Eq. (24), giving

$$\theta\left[\frac{d^2f}{d\lambda^2} + \left(\frac{df}{d\lambda}\right)^2\right] + \frac{d\theta}{d\lambda}\frac{df}{d\lambda} + \frac{K}{\theta} = 0\,. \tag{29}$$

Assuming

$$\left.\frac{d^2f/d\lambda^2}{(df/d\lambda)^2}\right|_H = 0\,, \tag{30}$$

then Eq. (29) is simply a quadratic equation for $df/d\lambda$, which can be solved to give

$$\frac{df}{d\lambda} \approx \frac{1}{\theta}\left(-\frac{1}{2}\frac{d\theta}{d\lambda} \pm iK^{1/2}\right)\,. \tag{31}$$

From Eq. (31) we get

$$\frac{d^2f/d\lambda^2}{(df/d\lambda)^2} \approx \pm\frac{i\,d\theta/d\lambda}{K^{1/2}} + \frac{1}{2}\frac{\theta d^2\theta/d\lambda^2 - (d\theta/d\lambda)^2}{K}\,, \tag{32}$$

which vanishes at the horizon, justifying the assumption of Eq. (30). So we find in the second case that the two linearly independent solutions of Eq. (24) have the following approximate form near the horizon,

$$\phi_\pm = \frac{\text{const}}{\theta^{1/2}} \times \exp\left(\pm iK^{1/2}\int d\lambda/\theta\right)\,. \tag{33}$$

Both solutions are singular at the horizon, again giving a contradiction with our initial assumptions. The conclusion of this somewhat lengthy analysis is that $A_t\,|_H \neq 0$ is not allowed, and thus by our earlier arguments, A_t must vanish identically outside the horizon. That is, a black hole cannot support an exterior massive vector-meson field, even when the mass is generated by spontaneous symmetry breaking.

III. THE GOLDSTONE MODEL ["MINIMAL" (ξ=0) CASE]

With $A_t \equiv 0$, Eqs. (1)–(12) of Sec. II describe the Goldstone model of a self-interacting scalar field, as generalized to curved space-time. We will now show that for this model in the "minimal" ($\xi = 0$) case, a further "no-hair" theorem can be proved, stating that $\phi \equiv \phi_\infty$ for all $r \geq r_H$. That is, outside the horizon the scalar field reduces to an unobservable constant, and [cf. Eq. (10a)] the scalar field stress-energy tensor vanishes identically. Our argument does not apply to the "conformal" ($\xi = \frac{1}{6}$) case, where the scalar field stress-energy tensor has a considerably more complicated structure than in the "minimal" case.

The argument proceeds from the scalar field equation, which with $A_t \equiv 0$ takes the form

$$\begin{aligned} &(p^{1/2}qr^2\phi')' = r^2p^{-1/2}\tfrac{1}{2}V(\phi)\,, \\ &V(\phi) = h(\phi^2-\phi_\infty^2)^2\,, \end{aligned} \tag{34}$$

and from the Einstein equations, which with $A_t \equiv 0$ may be rearranged to give

$$\begin{aligned} &p' = -16\pi rp(\phi')^2\,, \\ &(p^{1/2}qr)' = p^{-1/2} - 8\pi r^2p^{-1/2}V(\phi)\,. \end{aligned} \tag{35}$$

Multiplying Eq. (34) by ϕ' and integrating from r_H to ∞ gives, after use of Eq. (35) and an integration by parts,

$$\begin{aligned} 0 = \int_{r_H}^{\infty} dr\,[&\tfrac{1}{2}(\phi')^2p^{1/2}qr + \tfrac{1}{2}(\phi')^2p^{-1/2}r \\ &+ rp^{-1/2}V(\phi)] \\ &+ \tfrac{1}{2}[r^2p^{-1/2}V(\phi)]_H - [p^{1/2}qr^2\tfrac{1}{2}(\phi')^2]_H\,. \end{aligned} \tag{36}$$

Since all terms in Eq. (36) are non-negative except for the final one, we see that if $[p^{1/2}q(\phi')^2]_H$

$= 0$, then we can conclude that $\phi \equiv \phi_\infty$ for $r \geq r_H$, and the desired "no-hair" theorem follows.

To complete the proof, we must exclude the possibility $[p^{1/2}q(\phi')^2]_H \neq 0$. Just as in the preceding section, this is done by a local analysis in the vicinity of the horizon. We begin by noting that since $dq/d\lambda\,|_H \geq 0$, Eq. (20a) implies

$$\begin{aligned} G^{\hat{r}\hat{r}} + \frac{1}{r^2} &= \text{bounded} \\ &= \frac{p^{1/2}}{r}\left(\frac{dq}{d\lambda} + p^{1/2}\frac{q}{r}\right) \geq \frac{1}{r^2}\frac{1}{q}(p^{1/2}q)^2 \\ &\Rightarrow q \times \text{bounded} \geq (p^{1/2}q)^2 \\ &\Rightarrow p^{1/4}q^{1/2}|_H = 0 . \end{aligned} \tag{37}$$

Furthermore, since $T^{\hat{r}\hat{r}}\,|_H$ is bounded, and since both terms in $T^{\hat{r}\hat{r}}$ are positive semidefinite, we have that $p^{1/2}q^{1/2}\phi'\,|_H$ is bounded. Since the first equation in Eq. (35) implies that $dp/d(-r) \geq 0$, and since $p(\infty) = 1$, we have $p \geq 1$, which puts the boundedness of $p^{1/2}q^{1/2}\phi'$ into the form

$$p^{1/4}q^{1/2}\phi'|_H = \frac{p^{1/2}q^{1/2}\phi'\,|_H}{p^{1/4}\,|_H} = \text{bounded}. \tag{38}$$

Suppose now that $p^{1/4}q^{1/2}\phi'\,|_H = K \neq 0$. Then

$$\begin{aligned} \phi\,|_H = \text{bounded} &\Rightarrow \int_{r_H} dr\frac{d\phi}{dr} = \text{bounded} \\ &\Rightarrow \int_{r_H} \frac{dr}{a(r)} = \text{convergent} , \end{aligned} \tag{39}$$

with

$$a(r) \equiv p^{1/4}q^{1/2}r^2, \quad a\,|_H = 0 . \tag{40}$$

But on the other hand, the differential equation for ϕ in Eq. (34) gives

$$(p^{1/2}qr^2\phi')'\,|_H = \text{bounded}, \tag{41}$$

which on substituting $\phi' \approx K/(p^{1/4}q^{1/2})$ gives

$$\begin{aligned} a'\,|_H &= \text{bounded} \\ \Rightarrow \lim_{r \to r_H} \frac{a(r)}{r - r_H} &= \text{bounded} \\ \Rightarrow \int_{r_H} \frac{dr}{a(r)} &= \text{divergent} , \end{aligned} \tag{42}$$

in contradiction with Eq. (39). Hence we must have $K = 0$, which completes the proof.

IV. AN ABELIAN HIGGS ANALOG MODEL IN MINKOWSKI SPACE-TIME

As our final topic we briefly investigate a Minkowski space-time analog of the Abelian Higgs model analyzed in Sec. II. We consider a sphere of radius r_H, impenetrable to the Higgs field, and carrying charge Q, surrounded by the Higgs scalar medium. The differential equations and boundary conditions describing the time-independent behavior of this system are

$$\begin{aligned} &(r^2A_t')' = r^2 2e^2A_t\phi^2 , \\ &(r^2\phi')' = r^2[-e^2A_t^{\,2}\phi + 2h\phi(\phi^2 - \phi_\infty^{\,2})] , \\ &\phi(r_H) = 0, \quad A_t'(r_H) = -\frac{Q}{r_H^{\,2}} , \end{aligned} \tag{43}$$

which apart from the absence of the metric factors e^α, e^β have essentially the same structure as the system of equations analyzed in Sec. II. However, unlike the situation found in the general relativistic case, the Minkowski model of Eq. (43) has a nontrivial screened-charge solution.[9] To prove this, we consider the energy functional

$$E(r_H, Q) = 4\pi\int_{r_H}^\infty r^2dr[\tfrac{1}{2}(A_t')^2 + e^2A_t^{\,2}\phi^2 + h(\phi^2 - \phi_\infty^{\,2})^2] \tag{44}$$

and use the differential equation for A_t (the charge conservation constraint equation) and its associated boundary condition to write

$$A_t' = \frac{1}{r^2}\left[\int_{r_H}^r dr'r'^2 2e^2A_t(r')\phi^2(r') - Q\right] , \tag{45}$$

which when substituted into Eq. (45) gives the new functional,

$$^*E(r_H, Q) = 4\pi\int_{r_H}^\infty r^2dr\left\{\frac{1}{2r^4}\left[\int_{r_H}^r dr'r'^2 2e^2A_t(r')\phi^2(r') - Q\right]^2 + e^2A_t^{\,2}\phi^2 + h(\phi^2 - \phi_\infty^{\,2})^2\right\} . \tag{46}$$

Extremizing $^*E$ with respect to variations in A_t and ϕ [with an endpoint condition $\delta\phi(r_H) = 0$] is easily verified to lead to the differential equations of Eq. (43). Hence these equations describe the field configuration which minimizes the field energy, subject to the constraint that the inaccessible region $r \leq r_H$ contains total charge Q.

Since the functional $^*E$ is positive semidefinite, and since there is a nonempty class of functions A_t, ϕ for which $^*E$ is bounded from above, func-

tions A_t, ϕ which minimize $*E$ must exist, and thus the coupled equations in Eq. (43) have a solution.[10] Near $r=\infty$, the solution has the behavior

$$\begin{aligned} &\phi \approx \phi_\infty \\ &A_t \approx r^{-1} \exp[-r/(2e^2\phi_\infty^{\,2})^{1/2}], \end{aligned} \tag{47}$$

and as expected, the Higgs mechanism results in screening of the charge Q from view at infinity. The conclusion from this analysis is that the absence of screened-charge black-hole solutions in the general-relativistic case is a result of the stringent conditions for the existence of a horizon, not of the interacting field or spontaneous-symmetry-breaking aspects of the problem.

ACKNOWLEDGMENTS

We wish to thank P. Hohenberg and C. Teitelboim for useful conversations. This research was supported by the Department of Energy under Grant No. EY-76-S-02-2220.

---

[1]J. A. Wheeler, *Atti del Convegno Mendeleeviano* (Accademia delle Scienze di Torino, Accademia Nazionale dei Lincei, Torino-Roma, 1969).

[2]J. Bekenstein, Phys. Rev. D 5, 1239 (1972); 5, 2403 (1972); J. Hartle, *ibid.* 3, 2938 (1971); C. Teitelboim, *ibid.* 5, 2941 (1972).

[3]For a pedagogical review and references, see J. Bernstein, Rev. Mod. Phys. 46, 7 (1974).

[4]See, e.g., C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 14.

[5]Note that A_t is the time component on a coordinate basis.

[6]In the massless vector case, an exterior vector field is present precisely because the possibility $A_t|_H \neq 0$ can be realized.

[7]At this point in his analysis of the charged scalar-meson case, Bekenstein [J. D. Bekenstein, Phys. Rev. D 5, 1239 (1972)] introduces the assumption that the "charge per meson" which he defines by $(-j^\mu j_\mu)^{1/2}/\phi^2 \propto (A_t\phi^2 A^t\phi^2)^{1/2}/\phi^2 \propto e^{-\alpha}A_t$ is bounded at the horizon, which would imply the vanishing of $A_t|_H$ with no further detailed analysis. However, it is not clear to us that the requirement of boundedness at the horizon should apply to physical scalars formed as the *quotients* of other scalars, when the denominator is a physical quantity (such as ϕ) which can develop nodes. In our analysis, we only assume boundedness at the horizon of $F_{\mu\nu}F^{\mu\nu}$ and of $G_{\mu\nu}G^{\mu\nu} = (8\pi)^2 T_{\mu\nu}T^{\mu\nu}$. Since $G^{\hat\mu\hat\nu}$ is diagonal, boundedness of $G_{\mu\nu}G^{\mu\nu}$ implies that all components of $G^{\hat\mu\hat\nu}$ are individually bounded at the horizon.

[8]See C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Ref. 4), Chap. 25.

[9]Our variational argument for existence of a screened charge solution applies when the boundary condition $\phi(r_H)=0$ is generalized to $\phi(r_H)=\phi_H$, with ϕ_H any specified constant.

[10]The proof does not extend to the limit $r_H \rightarrow 0$ because a point charge has infinite Coulomb self-energy, as a result of which the functional $*E$ is not bounded from above. In the point charge case, an analysis of the indicial equation for ϕ around $r=0$ suggests that, for $(eQ)^2 < \frac{1}{4}$, there may be a solution which would behave as $\phi \sim \hat\phi r^\lambda$, $\lambda = -\frac{1}{2} + [\frac{1}{4} - (eQ)^2]^{1/2}$ near $r=0$, and which joins on to an exponentially decreasing asymptotic solution at $r=\infty$. However, we do not have an existence proof in this case.

PHYSICAL REVIEW LETTERS

VOLUME 44 16 JUNE 1980 NUMBER 24

Order-R Vacuum Action Functional in Scalar-Free Unified Theories with Spontaneous Scale Breaking

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 10 March 1980)

It is shown that in unified theories containing only fermions and gauge fields and in which scale invariance is spontaneously broken, radiative corrections induce an order-R term in the vacuum action functional which is uniquely determined by the flat-space-time theory.

PACS numbers: 04.60.+n, 11.30.Qc, 12.20.Hx

In recent papers Minkowski,[1] Zee,[1] and Smolin[2] have suggested that spontaneous scale-invariance breaking may play an important role in a fundamental theory of gravitation. The basic mechanism considered by both involves an action[3]

$$S=\int d^4x(-g)^{1/2}[\tfrac{1}{2}\epsilon\varphi^2R-V(\varphi^2)+\text{kinetic terms}+\text{other fields}], \tag{1}$$

with φ a scalar field. The potential $V(\varphi^2)$ is assumed to have a minimum away from $\varphi=0$,

$$V'(\kappa^2)=0,\qquad V''(\kappa^2)>0, \tag{2}$$

so that spontaneous symmetry breaking induces an effective gravitational action

$$S_{\text{grav}}=\int d^4x\,(-g)^{1/2}\tfrac{1}{2}\epsilon\kappa^2R, \tag{3}$$

with ϵ a free parameter. In this note I examine the analog of the above mechanism in unified theories which contain no fundamental scalar fields and in which scale invariance is spontaneously broken.[4] I show that in such theories the vacuum action functional contains an order-R term which is *explicitly calculable* in terms of the flat-space-time parameters of the theory. This result is basically an extension, to curved space-times, of the known fact that in such theories all mass ratios are explicitly calculable.

Consider a theory based on a scale-invariant classical Lagrangian, constructed from spin-$\frac{1}{2}$ fermion and spin-1 gauge fields, with the generally covariant renormalized matter action

$$\tilde{S}=\int d^4x\,(-g)^{1/2}(\mathcal{L}_{\text{matter}}+\text{counter terms}). \tag{4}$$

Because quantum effects induce nonlocal interactions with the space-time curvature, the vacuum action functional $\langle\tilde{S}\rangle_0$ *cannot* be related to its flat-space-time value by the equivalence principle. Instead, we have a formal decomposition

$$\langle\tilde{S}\rangle_0=\int d^4x\,(-g)^{1/2}\langle\tilde{\mathcal{L}}\rangle_0,$$
$$\langle\tilde{\mathcal{L}}\rangle_0=\sum_{n=0}^{\infty}\kappa^{4-2n}\langle\tilde{\mathcal{L}}\rangle_0^{(2n)}$$
$$=\kappa^4\langle\tilde{\mathcal{L}}\rangle_0^{(0)}+\kappa^2\langle\tilde{\mathcal{L}}\rangle_0^{(2)}+\langle\tilde{\mathcal{L}}\rangle_0^{(4)}+\dots, \tag{5}$$

with κ the unification mass of the flat-space-time theory and with $\langle\tilde{\mathcal{L}}\rangle_0^{(2n)}$ homogeneous of degree $2n$ in derivatives ∂_x acting on the metric. Because the curvature scalar R is the only Lorentz

scalar of order $(\partial_x)^2$, the second term in Eq. (5) has the form[5]

$$\langle\tilde{\mathcal{L}}\rangle_0^{(2)} = \beta R, \tag{6}$$

with β in general nonzero. According to a criterion of Weinberg,[6] the coefficient β will be calculable in the flat-space-time field theory, provided that there are *no possible* Lagrangian counter terms which contribute to this term. The only relevant counter terms of the general form

$$\Delta\mathcal{L} = \mathcal{O}_2 R, \tag{7}$$

with $\mathcal{O}_2$ a gauge-invariant operator with canonical dimension 2. However, in a theory with no fundamental scalars, and with spontaneous breaking of scale invariance (and hence no bare-mass parameters), the only dimension-2 operators are of the form $b_\mu{}^a b^{\mu a}$, with $b_\mu{}^a$ a gauge potential. But such operators are not gauge invariant, and hence no counter terms of the form of Eq. (7) are possible. Therefore, in the theories under consideration, β is finite and calculable (as opposed to merely renormalizable).

Following[7] Sakharov[8] and Klein,[9] it is tempting to regard the $\kappa^2\beta R$ term in Eq. (5) as the entire gravitational action, rather than as just an additional finite contribution to the gravitational action. This interpretation is clearly justified if the unified matter theory predicts the correct sign and magnitude of the Einstein action *and* if the virtual integrations contributing to β are dynamically cut off at energies well below the Planck mass. If the virtual integrations extend beyond the Planck mass, then use of the semiclassical, background metric analysis given above requires further justification or corrections, involving an analysis of possible quantum gravity effects.[10]

Added notes.—Calculations by Hasslacher and Mottola,[11] Mottola,[11] and Zee[11] in models obeying the premises of this note all give a nonvanishing induced order-R term, and show that the sign can correspond to attractive gravity.

Guo has brought to my attention a number of further references on R^2-type gravity Lagrangians.[12] In particular, it is known that a $C_{\mu\nu\lambda\sigma}\times C^{\mu\nu\lambda\sigma}$ gravity theory is renormalizable, but has a dipole ghost. Hence the extended matter-gravity theories discussed in Ref. 10 of this note are renormalizable; they could also be unitary (by the Lee-Wick[13] mechanism) if scale-symmetry breakdown causes the dipole ghost to split into a single positive-residue graviton pole at $k^2=0$, and a pair of complex-conjugate unstable ghost poles at $k^2=M\pm i\Gamma$ (with M and Γ of order the unification mass). Detailed dynamical studies of the extended matter-gravity theories will be needed to settle the unitarity issue. Tomboulis[14] has given an interesting model with a dynamical Lee-Wick mechanism, and I wish to thank him for a discussion about this point.

Because dimension-4 and dimension-0 operators are always available (e.g., $\mathcal{L}_{\rm matter}$ and the gauge-field bare coupling, respectively), the arguments of this note do not apply to the order-(0) or -(4) terms in Eq. (5). Hence, even in scalar-free theories with spontaneous scale-invariance breaking, the cosmological constant contains renormalizable infinities. An additional symmetry, very likely relating the boson and fermion sectors of the theory, will be needed to give a calculable cosmological constant. L. S. Brown and J. C. Collins point out that because dimension-0 operators are available, the induced R^2 term in the vacuum action can become divergent in high loop order; if this happens, a quadratic gravitational Lagrangian must include an R^2 term in addition to the term $C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma}$ discussed above.

I wish to thank L. S. Brown, J. C. Collins, D. J. Gross, H.-Y. Guo, B. Hasslacher, E. Mottola, M. J. Perry, and T. Tomboulis for helpful comments. This work was supported by the U. S. Department of Energy under Grant No. DE-AC02-76ER02220.

[1]P. Minkowski, Phys. Lett. 71B, 419 (1977); A. Zee, Phys. Rev. Lett. 42, 417 (1979).

[2]L. Smolin, Nucl. Phys. B160, 253 (1979). Smolin discusses a conformal-invariant scalar-vector theory, where the equation governing the classical minimum is $(\partial/\partial\varphi - 4/\varphi)V = 0$, rather then simply $V' = 0$ as in Minkowski's or Zee's model.

[3]I follow the conventions of C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[4]See, e.g., L. Susskind, Phys. Rev. D 20, 2619 (1979); S. Weinberg, Phys. Rev. D 13, 974 (1976), and D 19, 1277 (1979); or the $m_0 = 0$ version of S. L. Adler, Phys. Lett. 86B, 203 (1979), and "Quaternionic Chromodynamics as a Theory of Composite Quarks and Leptons," Phys. Rev. D (to be published).

[5]The term $\langle\mathcal{L}\rangle_0^{(4)}$, in addition to local contributions proportional to R^2, ... has nonlocal contributions arising from the effects of massless fields. See, for example, L. S. Brown and J. P. Cassidy, Phys. Rev. D 16, 1712 (1977). I am assuming that such nonlocal

terms do not appear in $\langle \mathfrak{L} \rangle_0^{(2)}$, but the presence of such terms would not alter the argument for the calculability of β. In noncompact manifolds there can be Lorentz scalars other than R which contribute (I wish to thank S. M. Christensen for this remark), but again these do not alter the argument given for the βR term.

[6]S. Weinberg, Phys. Rev. Lett. 29, 388 (1972). For a recent pedagogical review of the renormalization algorithm implicit in the calculability criterion, see L. S. Brown, "Dimensional Renormalization of Composite Operators in Scalar Field Theory" (unpublished).

[7]See C. W. Misner, K. S. Thorne, and J. A. Wheeler, Ref. 3, pp. 426–428.

[8]A. D. Sakharov, Dokl. Akad. Nauk. SSSR 177, 70 (1967) [Sov. Phys. Dokl. 12, 1040 (1968)].

[9]O. Klein, Phys. Scr. 9, 69 (1974).

[10]One possibility is that the metric is not a quantum variable, but is a classical dynamical variable governed by the Euler-Lagrange equations, in which case the background metric analysis given in the text is exact. Another possibility consistent with the viewpoint of the text is that the metric is a quantum variable, with dynamics governed by a scale-invariant fundamental Lagrangian (see L. Smolin, Ref. 2). The only generalization of Eq. (4) to include a scale-invariant gravitational action is $\tilde{S} = \int d^4x\, (-g)^{1/2}\, (\mathfrak{L}_{\mathrm{matter}} + \delta C_{\mu\nu\lambda\sigma} C^{\mu\nu\lambda\sigma} + \text{counterterms})$, with $C_{\mu\nu\lambda\sigma}$ the Weyl tensor, and with δ a dimensionless coupling constant. Recent work of Stelle [K. S. Stelle, Phys. Rev. D 16, 953 (1977)] on quadratic gravitational actions suggests that this extended theory should still be renormalizable. The $\kappa^2\beta R$ term in Eq. (5) would still be calculable, even with quantum gravitational effects taken into account, but the β value calculated from the flat-spacetime matter theory would be subject to a finite, δ-dependent renormalization. This renormalization could be important if the virtual integrations contributing to β extend to energies beyond the Planck mass.

[11]B. Hasslacher and E. Mottola, unpublished; E. Mottola, unpublished; A. Zee, unpublished.

[12]D. E. Neville, Phys. Rev. D 18, 3535 (1978), and unpublished; E. Sezgin and P. van Nieuwenhuizen, unpublished; H.-Y. Guo *et al.*, unpublished.

[13]T. D. Lee and G. C. Wick, Nucl. Phys. B9, 209 (1969); B10, 1 (1964); Phys. Rev. D 2, 1033 (1970).

[14]T. Tomboulis, Phys. Lett. 70B, 361 (1977).

Volume 95B, number 2 PHYSICS LETTERS 22 September 1980

A FORMULA FOR THE INDUCED GRAVITATIONAL CONSTANT

Stephen L. ADLER
The Institute for Advanced Study, Princeton, NJ 08540, USA

Received 27 May 1980

I derive a formula for the induced gravitational constant in unified theories with dynamical scale-invariance breakdown.

In a recent note [1] I gave a simple argument showing that in unified theories with dynamical scale-invariance breakdown, radiative corrections in curved spacetime induce an order-R term in the vacuum action functional which is uniquely determined by the flat spacetime theory. Subsequent calculations by Hasslacher and Mottola [2], Zee [3] and Mottola [4] in models obeying the premises of ref. [1] all give a non-vanishing induced gravitational constant, and show [2] that its sign can correspond to attractive gravity. Their calculations also suggest that it should be possible to do the curved spacetime manipulations in a formal way, giving an explicit formula for the induced gravitational constant involving flat spacetime quantities only. I derive such a formula below; it leads to a less abstract derivation of the basic finiteness theorem of ref. [1], and should provide a useful starting point for dynamical calculations.

Consider a theory based on a scale-invariant classical lagrangian, constructed from spin-1/2 fermions and spin-1 gauge fields, with the generally covariant renormalized matter action

$$\widetilde{S} = \int d^4x \sqrt{-g}\, \widetilde{\mathcal{L}}_{\text{matter}}, \qquad \widetilde{\mathcal{L}}_{\text{matter}} = \mathcal{L}_{\text{matter}} + \text{counter-terms}\,. \tag{1}$$

Because quantum effects are non-local, the vacuum action functional $\langle\widetilde{S}\rangle_0$ in curved spacetime cannot be related to its flat spacetime value by the equivalence principle. Instead, we have a formal decomposition [‡1]

$$\langle\widetilde{S}\rangle_0 = \int d^4x \sqrt{-g}\, \langle\widetilde{\mathcal{L}}\rangle_0\,, \quad \langle\widetilde{\mathcal{L}}\rangle_0 = \langle\widetilde{\mathcal{L}}\rangle_0^{\text{flat spacetime}} + (16\pi G_{\text{ind}})^{-1}R + \text{terms involving higher metric derivatives,} \tag{2}$$

with G_{ind} the induced gravitational constant. I use sign conventions [‡2] in which a positive induced gravitational constant corresponds to attractive gravity. Defining the renormalized matter stress–energy tensor

$$\widetilde{T}^{\mu\nu} = 2(-g)^{-1/2}\frac{\delta}{\delta g_{\mu\nu}}\left[(-g)^{1/2}\widetilde{\mathcal{L}}_{\text{matter}}\right], \tag{3}$$

Eq. (2) can be rewritten as a formula for $\langle\widetilde{T}_\mu{}^\mu\rangle_0$,

$$\langle\widetilde{T}_\mu{}^\mu\rangle_0 = \langle\widetilde{T}_\mu{}^\mu\rangle_0^{\text{flat spacetime}} + (8\pi G_{\text{ind}})^{-1}R + \text{terms involving higher metric derivatives}\,. \tag{4}$$

To get a formula for G_{ind}, it suffices to calculate the change in $\langle\widetilde{T}_\mu{}^\mu\rangle_0$ induced by spacetime curvature, in the special case of a conformally flat, constant curvature spacetime. For a general lagrangian variation, we have [‡3]

‡1 In ref. [1], I denoted $(16\pi G_{\text{ind}})^{-1}$ by $\beta\kappa^2$.
‡2 I use the conventions of Misner et al. [5].
‡3 Eq. (5) is obtained from eq. (17.22) of Bjorken and Drell [6] by making the replacements $\varphi_{\text{in}} \to \widetilde{T}_\mu{}^\mu{}_\lambda$, $-\mathcal{H}_I(t) \to \int d^3x \sqrt{-g} \times \delta\mathcal{L}$, and is independent of metric conventions. Neglect of the intrinsic metric dependence [I take $\delta\widetilde{T}_\mu{}^\mu(0) = 0$] is allowed in a calculation to order-R, since this terms first contributes in order-R^2.

Volume 95B, number 2 PHYSICS LETTERS 22 September 1980

$$\delta\langle\widetilde{T}_\mu{}^\mu(0)\rangle_0 = \mathrm{i}\int \mathrm{d}^4x\ \{\langle T(\delta[(-g)^{1/2}\ \widetilde{\mathcal{L}}(x)]\widetilde{T}_\mu{}^\mu(0))\rangle_0 - \langle\delta[(-g)^{1/2}\widetilde{\mathcal{L}}(x)]\rangle_0\langle\widetilde{T}_\mu{}^\mu(0)\rangle\}$$

$$= \mathrm{i}\int \mathrm{d}^4x\ \langle T(\delta[(-g)^{1/2}\widetilde{\mathcal{L}}(x)]\widetilde{T}_\mu{}^\mu(0))\rangle_{0,\,\mathrm{connected}}\ . \tag{5}$$

Taking the metric to have the form near x of [‡4]

$$g_{\mu\nu}(x) = \eta_{\mu\nu}(1 - \tfrac{1}{24}Rx^2 + \ldots)\ , \tag{6a}$$

and taking $\delta\mathcal{L}$ to be the lagrangian change induced by the metric variation

$$\delta g_{\mu\nu}(x) = -\eta_{\mu\nu}\cdot\tfrac{1}{24}Rx^2\ , \tag{6b}$$

we immediately get [by a second use of eq. (3)]

$$(8\pi G_{\mathrm{ind}})^{-1}R = \delta\langle\widetilde{T}_\mu{}^\mu(0)\rangle_0 = \mathrm{i}\int \mathrm{d}^4x\ (-g)^{1/2}(-\tfrac{1}{24}Rx^2)\langle T(\tfrac{1}{2}\widetilde{T}_\lambda{}^\lambda(x)\widetilde{T}_\mu{}^\mu(0))\rangle_{0,\,\mathrm{connected}}\ . \tag{7}$$

Dividing by R, and taking the limit of flat spacetime, gives the desired formula

$$(16\pi G_{\mathrm{ind}})^{-1} = \frac{\mathrm{i}}{96}\int \mathrm{d}^4x\ [(x^0)^2 - \boldsymbol{x}^2]\,\langle T(\widetilde{T}_\lambda{}^\lambda(x)\widetilde{T}_\mu{}^\mu(0))\rangle^{\text{flat spacetime}}_{0,\,\text{connected}}\ . \tag{8}$$

To study the ultraviolet convergence properties of eq. (8), it is necessary to regard eq. (8) as the dimensional continuation limit

$$(16\pi G_{\mathrm{ind}})^{-1} = \frac{\mathrm{i}}{96}\lim_{n\to 4}\int \mathrm{d}^nx\ [(x^0)^2 - \boldsymbol{x}^2]\langle T(\widetilde{T}_\lambda{}^\lambda(x)\widetilde{T}_\mu{}^\mu(0))\rangle^{\text{flat spacetime}}_{0,\,\text{connected}}\ . \tag{9}$$

From a perturbative point of view [7], the only way poles at $n = 4$ can appear is from terms of the form

$$\langle T(\widetilde{T}_\lambda{}^\lambda(x)\widetilde{T}_\mu{}^\mu(0))\rangle^{\text{flat spacetime}}_{0,\,\text{connected}} = \ldots + \langle\mathcal{O}_2\rangle^{\text{flat spacetime}}_{0,\,\text{connected}} \times (x^2)^{-3}\times \text{logarithms} + \ldots\ , \tag{10}$$

in the perturbative operator product expansion of the T-product, with $\mathcal{O}_2$ a gauge-invariant operator of canonical dimension 2. But the hypothesis of dynamical scale invariance breakdown (vanishing bare masses, no scalar fields) excludes the presence of such operators, and so the limit of eq. (9) as $n\to 4$ is finite. From a nonperturbative point of view, the $(x^2)^{-3}$ term in the expansion of eq. (10) is altered, in theories with dynamical scale breaking [‡5], to either of the forms

$$(x^2)^{-3+\gamma}\ ,\quad \gamma > 0\ ;\qquad (x^2)^{-3}(\log x^2)^{-\delta}\ ,\quad \delta > 1\ , \tag{11}$$

for both of which the limit of eq. (9) exists as $n\to 4$. Note that although the classical lagrangian of eq. (4) is conformally invariant, dynamical scale breaking introduces a mass scale into the theory, and so low energy matrix elements of $\widetilde{T}_\mu{}^\mu(x)$ will be non-vanishing. Hence eq. (8) gives an ultraviolet-convergent, and in general non-vanishing, induced inverse gravitational constant.

I have assumed up to this point that eq. (8) is infrared finite, as is true in the calculation of Zee [3]. In the instanton gas model examined by Hasslacher and Mottola [2], the leading curvature term in $\langle\widetilde{\mathcal{L}}\rangle_0$ is of the form $R\log(1/R)$, indicating that eq. (8) is logarithmically divergent at $x = \infty$. The divergence arises from expanding the

‡4 The local expansion of a general conformally flat metric is

$$g_{\mu\nu}(x) = \eta_{\mu\nu}(1 + \tfrac{1}{4}E_{\xi\eta}x^\xi x^\eta + \ldots),\quad E_{\xi\eta} = -2R_{\xi\eta} + \tfrac{1}{3}\eta_{\xi\eta}R\ .$$

Eq. (6a) results from making the specialization $R_{\xi\eta} = \frac{1}{4}Rg_{\xi\eta}$.

‡5 See Pagels [8] for a review of models of dynamically broken gauge theories.

conformal factor in a power series in x inside the x integral [cf. eq. (6a)]; the legality of this expansion is not guaranteed by the ultraviolet-finiteness criteria of ref. [1].

One can attempt to carry the argument one step further, by using dispersion relations to put eq. (8) in spectral form. Defining

$$\psi(q^2) \equiv \int d^4x\, e^{iq\cdot x}(-i)\langle T(\widetilde{T}_\lambda{}^\lambda(x)\widetilde{T}_\mu{}^\mu(0))\rangle_{0,\text{ connected}}^{\text{flat spacetime}} , \tag{12}$$

and assuming that $\psi(q^2) - \psi(0)$ satisfies an unsubtracted dispersion relation, a simple calculation [‡6] gives

$$(16\pi G_{\text{ind}})^{-1} = -\frac{1}{12}\int_0^\infty d\sigma^2\, \rho(\sigma^2)/\sigma^4 , \qquad \rho(-q^2) = (2\pi)^3 \sum_n \delta^4(p_n - q)|\langle 0|\widetilde{T}_\mu{}^\mu(0)|n\rangle|^2 . \tag{13}$$

Since in canonical gauges the Hilbert space metric is positive definite, eq. (13) implies that G_{ind} is negative. However, both eq. (13) and this conclusion about the sign of G_{ind} are false. The reason is that in gauge theories $\widetilde{T}_\mu{}^\mu$ contains [9] a trace anomaly term proportional to $\beta(g^2)N(F^a_{\mu\nu}F^{\mu\nu a})$, which makes the spectral function behave asymptotically as $\rho \sim \sigma^4 \times$ logarithms, invalidating the unsubtracted dispersion relation assumption needed to derive eq. (13). Although the trace anomaly gives rise to a singular term proportional to $\beta^2(x^2)^{-4} \times$ logarithms in the operator product expansion of eq. (10), the contribution of this term to eq. (9) is well behaved in the $n \to 4$ dimensional limit.

I wish to thank B. Hasslacher, L. Brown, E. Mottola and A. Zee for numerous discussions, and E. Mottola for checking the calculation. A. Zee has independently obtained some of the formulas given above, and suggested the role of σ^4 terms in ρ in the breakdown of the spectral analysis. This work was supported by the Department of Energy under Grant No. DE-AC02-76ER02220.

[‡6] Eqs. (14) and (15) follow immediately from eqs. (16.33), (16.27) and appendix C of Bjorken and Drell [6].

References

[1] S.L. Adler, Phys. Rev. Lett., to be published.
[2] B. Hasslacher and E. Mottola, Phys. Lett. 95B (1980) 237.
[3] A. Zee, submitted to Phys. Rev. D.
[4] E. Mottola, unpublished.
[5] C.W. Misner, K.S. Thorne and J.A. Wheeler, Gravitation (Freeman, San Francisco, 1973).
[6] J.D. Bjorken and S.D. Drell, Relativistic quantum fields (McGraw-Hill, New York, 1965).
[7] W. Zimmermann, Local operator products and renormalization in quantum field theory, in: Lectures on elementary particles and quantum field theory, Vol. 1 (1970 Brandeis lectures), eds. S. Deser, M. Grisaru and H. Pendleton (MIT Press, Cambridge, MA, 1970).
[8] H. Pagels, Phys. Rev. D, to be published.
[9] S.L. Adler, J.C. Collins and A. Duncan, Phys. Rev. D15 (1977) 1712;
J.C. Collins, A. Duncan and S.D. Jogkelar, Phys. Rev. D16 (1977) 438.

Einstein gravity as a symmetry-breaking effect in quantum field theory*

Stephen L. Adler

The Institute for Advanced Study, Princeton, New Jersey 08540

This article gives a pedagogical review of recent work in which the Einstein-Hilbert gravitational action is obtained as a symmetry-breaking effect in quantum field theory. Particular emphasis is placed on the case of renormalizable field theories with dynamical scale-invariance breaking, in which the induced gravitational effective action is finite and calculable. A functional integral formulation is used throughout, and a detailed analysis is given of the role of dimensional regularization in extracting finite answers from formally quadratically divergent integrals. Expressions are derived for the induced gravitational constant and for the induced cosmological constant in quantized matter theories on a background manifold, and a strategy is outlined for computing the induced constants in the case of an SU(n) gauge theory. By use of the background field method, the formalism is extended to the case in which the metric is also quantized, yielding a derivation of the semiclassical Einstein equations as an approximation to quantum gravity, as well as general formulas for the induced (or renormalized) gravitational and cosmological constants.

CONTENTS

I. INTRODUCTION

In the conventional formulation of general relativity, gravitation is described by rewriting the matter action in generally covariant form, and by adding to it the Einstein-Hilbert gravitational action

$$S_{\mathrm{grav}}=\frac{1}{16\pi G}\int d^4x\,\sqrt{-g}\,(R-2\Lambda)\,, \tag{1.1}$$

with G Newton's constant and R the curvature scalar, and with the cosmological constant Λ taken to be zero. The total action is then treated as a classical variational principle, to be extremized with respect to variations of the c-number metric $g_{\mu\nu}$. As discussed in the survey articles in Hawking and Israel (1979), the theory in this form accounts very well for all astronomical gravitational phenomena and has a structure which is understood in considerable theoretical detail. On the other hand, when treated as a fundamental quantum action, Eq. (1.1) leads to a nonrenormalizable quantum field theory. This problem has long been known, and has stimulated much theoretical effort aimed at achieving a satisfactory quantization of the Einstein-Hilbert action or its supergravity extensions. [For reviews of the current status of these approaches, see Hawking and Israel (1979) and Van Nieuwenhuizen (1981).]

An entirely different approach to quantum gravity derives from work by Zel'dovich (1967) and Sakharov (1967) on induced quantum effects. Zel'dovich studied the effect of vacuum quantum fluctuations on the

*This article is based on the Eleventh Lecture in the Alfred Schild Memorial Series, given at The University of Texas at Austin, April 8—16, 1981.

cosmological constant; extending this idea, Sakharov proposed that Eq. (1.1) is not a fundamental microscopic action, but rather is an effective action induced by vacuum quantum structure (see also Klein, 1974). To quote the two key sentences from Sakharov's paper, "The presence of the action (1) [Eq. (1.1)] leads to a metrical elasticity of space, i.e., to generalized forces which oppose the curving of space. (¶) Here we consider the hypothesis which identifies the action (1) with the change in the action of quantum fluctuations of the vacuum if space is curved." Sakharov's proposal attracted attention from the outset (see Misner *et al.*, 1970), but further progress was hampered by the fact that in the free field models for which he made his estimates, the induced gravitational constant $G_{\rm ind}^{-1}$ is given by integrals which contain both quadratic and logarithmic divergences. It is only in the last few years that the technology of quantum field theory has advanced to the point where one can systematically study induced quantum effects in interacting field theories. These advances, and their application to induced Einstein gravity, are the subject matter of this review.

Since the topics discussed below span the areas of high-energy physics and relativity, in which different notational conventions are generally used, I have adopted the following compromise with respect to notation. I use microscopic units throughout,

$$\hbar=c=1\ , \tag{1.2}$$

so the only dimensional quantity is mass=(length)$^{-1}$. The coordinates x^μ are taken to have the dimension (length)^1, making the metric $g_{\mu\nu}$ dimensionless. In all flat space-time examples and discussions, I use the $+---$ signature convention of Bjorken and Drell (1965), while in all expressions which involve a curved manifold I follow the $-+++$ convention of Misner *et al.*, (1970). In the few places where it is necessary to change from one convention to the other, I will explicitly call attention to the transition.

II. FIELD THEORY PRELIMINARIES

A. Actions and canonical dimension accounting

The functional integral formulation of quantum field theory (see Abers and Lee, 1973) expresses transition amplitudes in the form

$$\begin{aligned} Z&=\int d\{\phi\}e^{iS[\{\phi\}]}\ ,\\ S[\{\phi\}]&=\int d^4x\,\mathscr{L}[\{\phi\}]\ , \end{aligned} \tag{2.1}$$

with $\{\phi\}$ the set of fields present, S the action, and $\mathscr{L}$ the Lagrangian (or action) density. Since the argument of an exponential or a logarithm must be dimensionless, in the conventional accounting of canonical dimension in which

$$\dim[\text{length}]=-1,\quad \dim[\text{mass}]=+1\ , \tag{2.2}$$

we have

$$\begin{aligned} &\dim[S]=0,\quad \dim[d^4x]=-4\\ &\qquad\Longrightarrow \dim[\mathscr{L}]=4\ . \end{aligned} \tag{2.3}$$

From Eq. (2.3) we can infer the canonical dimensionality of the fields and parameters from which elementary renormalizable matter theories are constructed. For a scalar φ^4 field theory we have

$$\mathscr{L}=\tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi-\tfrac{1}{2}m_0^2\varphi^2-\tfrac{1}{4}\lambda_0\varphi^4\ , \tag{2.4}$$

with m_0 the bare mass and λ_0 the bare coupling, and with

$$\begin{aligned} &\dim[\partial_\mu\equiv\partial/\partial x^\mu]=1\Longrightarrow\dim[\varphi]=1\ ,\\ &\dim[m_0]=1\ ,\\ &\dim[\lambda_0]=0\ . \end{aligned} \tag{2.5}$$

For a spin-1 Abelian gauge field (the photon) we have

$$\begin{aligned} \mathscr{L}&=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}\ ,\\ F_{\mu\nu}&=\partial_\nu A_\mu-\partial_\mu A_\nu\ , \end{aligned} \tag{2.6}$$

with

$$\begin{aligned} &\dim[F_{\mu\nu}]=2\ ,\\ &\dim[A_\mu]=1\ , \end{aligned} \tag{2.7}$$

while for a spin-$\frac{1}{2}$ Dirac field we have

$$\mathscr{L}=\bar\psi(i\gamma^\mu\partial_\mu-m_0)\psi\ , \tag{2.8}$$

with

$$\dim[\psi]=\tfrac{3}{2}\ . \tag{2.9}$$

Minimal coupling of the photon to a Dirac field or a complex scalar field with bare charge e_0 yields the Lagrangian densities for quantum electrodynamics,

$$\begin{aligned} \mathscr{L}_{\rm QED1/2}&=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+\bar\psi(i\gamma^\mu D_\mu-m_0)\psi\ ,\\ \mathscr{L}_{\rm QED0}&=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+\tfrac{1}{2}|D_\mu\varphi|^2-\tfrac{1}{2}m_0^2|\varphi|^2-\tfrac{1}{4}\lambda_0|\varphi|^4\ ,\\ D_\mu&=\partial_\mu+ie_0A_\mu\ ,\\ &\Longrightarrow\dim[e_0]=0\ . \end{aligned} \tag{2.10}$$

For a spin-1 non-Abelian gauge field (the massless gauge gluon) we have

$$\begin{aligned} \mathscr{L}&=-\tfrac{1}{4}F_{\mu\nu}^iF^{i\mu\nu}\ ,\\ F_{\mu\nu}^i&=\partial_\nu A_\mu^i-\partial_\mu A_\nu^i+g_0f^{ijk}A_\mu^jA_\nu^k\ , \end{aligned} \tag{2.11}$$

with i the internal symmetry index, f^{ijk} the group structure constants, and g_0 the bare coupling constant. Minimal coupling of the gauge field to a Dirac field in the fundamental representation (with representation matrices $\frac{1}{2}\lambda^i$) gives the basic Lagrangian density for quantum chromodynamics,

$$\mathscr{L}_{\rm QCD}=-\tfrac{1}{4}F^{l}_{\mu\nu}F^{l\mu\nu}+\bar{\psi}(i\gamma^{\mu}D_{\mu}-m_0)\psi\ ,$$
$$D_{\mu}=\partial_{\mu}+ig_0\tfrac{1}{2}\lambda^{l}A^{l}_{\mu}\ , \qquad (2.12)$$

from which we infer the dimensional assignments

$$\dim[F^{l}_{\mu\nu}]=2\ ,$$
$$\dim[A^{l}_{\mu}]=1,\ \dim[\psi]=\tfrac{3}{2}\ ,$$
$$\dim[g_0]=0\ . \qquad (2.13)$$

All of the field theory models currently under study as candidates for unified matter theories [for reviews see Marciano and Pagels (1978); Fritzsch and Minkowski (1981)] are combinations of scalar, Dirac, and gauge field action building blocks of the basic types enumerated above. The characterizing feature of all such renormalizable actions is that their coupling constants (λ_0, e_0, g_0 above) have canonical dimension zero.

B. Effective actions

Consider now a renormalizable field theory with action

$$S[\{\phi^L\},\{\phi^H\}]=\int d^4x\,\mathscr{L}[\{\phi^L\},\{\phi^H\}]\ , \qquad (2.14)$$

where $\{\phi^L\}$ are "light" field components whose dynamics we directly observe, while $\{\phi^H\}$ are "heavy" field components which influence the dynamics of the light components but are not directly observable. The $\{\phi^H\}$ can in general include fields with high physical masses and high-momentum[1] components of fields with light physical masses. Since the $\{\phi^H\}$ are hidden from view, it is convenient to rewrite the functional integral of Eq. (2.1) in the following form,

$$Z=\int d\{\phi^L\}d\{\phi^H\}e^{iS[\{\phi^L\},\{\phi^H\}]}$$
$$=\int d\{\phi^L\}e^{iS_{\rm eff}[\{\phi^L\}]}\ , \qquad (2.15)$$

where the *effective action* $S_{\rm eff}[\{\phi^L\}]$ for the light fields is defined through

$$e^{iS_{\rm eff}[\{\phi^L\}]}=\int d\{\phi^H\}e^{iS[\{\phi^L\},\{\phi^H\}]}\ . \qquad (2.16)$$

Clearly, the effective action, if exactly known, would give a complete description of the dynamics of the fields $\{\phi^L\}$. In practice, one usually works with only an approximation to $S_{\rm eff}$, obtained by keeping leading terms in an expansion in a small parameter. Some examples of commonly used effective actions are as follows.

1. The Heisenberg-Euler (1936) effective action in quantum electrodynamics (see also Schwinger, 1951).

Integrating out the electron fields in quantum electrodynamics gives an effective action describing the nonlinear interactions of photons,

$$e^{iS_{\rm eff}[F_{\mu\nu}]}=\int d\{\psi,\bar{\psi}\}e^{iS_{\rm QED1/2}[\{F_{\mu\nu},\psi,\bar{\psi}\}]}\ . \qquad (2.17)$$

For field strengths which are slowly varying over an electron Compton wavelength, $S_{\rm eff}$ can be approximated by taking $F_{\mu\nu}$=constant, which gives a problem which can be solved in closed form. For weak, slowly varying fields (on a scale of an electron Compton wavelength), $S_{\rm eff}$ can be approximated by the first two terms in a series expansion

$$S_{\rm eff}[F_{\mu\nu}]=\int d^4x\,\mathscr{L}_{\rm eff}\ ,$$
$$\mathscr{L}_{\rm eff}=\tfrac{1}{2}(E^2-H^2)+\frac{2\alpha^2}{45m^4}[(E^2-H^2)^2+7(E\cdot H)^2]+\cdots\ , \qquad (2.18)$$

with E,H the electric and magnetic fields, α the fine-structure constant, and m the electron mass. If interpreted as a fundamental action and used (or, rather, misused) beyond the tree-approximation level, Eq. (2.18) would yield a nonrenormalizable perturbation expansion in powers of the dimensional effective coupling $2\alpha^2/45m^4$.

2. The four-fermion effective action approximation to the Weinberg (1967)-Salam (1968) weak interaction theory

At center-of-mass energies well below 100 GeV, the weak interactions are described by a current-current four-fermion effective action

$$S_{\rm eff}[\{\text{fermions}\}]=\int d^4x(\mathscr{L}^{\rm charged}_{\rm eff}+\mathscr{L}^{\rm neutral}_{\rm eff})\ ,$$
$$\mathscr{L}^{\rm charged}_{\rm eff}=\frac{1}{\sqrt{2}}G_F(j^{\lambda}_{\rm ch}+J^{\lambda}_{\rm ch})(j^{\dagger}_{{\rm ch}\lambda}+J^{\dagger}_{{\rm ch}\lambda})\ ,$$
$$\mathscr{L}^{\rm neutral}_{\rm eff}=\frac{1}{\sqrt{2}}G_F(j^{\lambda}_{n}+J^{\lambda}_{n})(j_{n\lambda}+J_{n\lambda})\ ,$$
$$j^{\lambda}_{\rm ch}=\bar{e}\gamma^{\lambda}(1-\gamma_5)\nu_e+\mu,\tau\ \text{terms}\ ,$$
$$J^{\lambda}_{\rm ch}=\bar{u}\gamma_{\lambda}(1-\gamma_5)(d\cos\theta_C+s\sin\theta_C)+\text{charm terms}\ ,$$
$$j^{\lambda}_{n}=-\tfrac{1}{2}\bar{e}\gamma^{\lambda}(1-\gamma_5)e+\tfrac{1}{2}\bar{\nu}_e\gamma^{\lambda}(1-\gamma_5)\nu_e+2\sin^2\theta_W\bar{e}\gamma^{\lambda}e+\mu,\tau\ \text{terms}\ ,$$
$$J^{\lambda}_{n}=\tfrac{1}{2}\bar{u}\gamma^{\lambda}(1-\gamma_5)u-\tfrac{1}{2}\bar{d}\gamma^{\lambda}(1-\gamma_5)d-2\sin^2\theta_W(\tfrac{2}{3}\bar{u}\gamma^{\lambda}u-\tfrac{1}{3}\bar{d}\gamma^{\lambda}d)+\text{strange, charm terms}\ , \qquad (2.19)$$

with e,ν_e,u,d,s the electron, electron neutrino, and up, down, and strange quark fields, respectively, and with θ_C and θ_W the Cabibbo and Weinberg angles.[2] Since the fermion fields have dimension $\frac{3}{2}$, the Fermi constant G_F has dimension -2, and has the empirical value

$$G_F\approx\frac{1.023\times10^{-5}}{m^2_{\rm proton}}\ . \qquad (2.20)$$

[1] I wish to thank B. Holdom for suggesting the inclusion of a momentum criterion in the definition, as a way of automatically including renormalization effects arising from overlapping divergences. For recent discussions of effective actions, see Weinberg (1980a) and Ovrut and Schnitzer (1980,1981).

[2] For a recent review of the phenomenology of the Weinberg-Salam model, see Kim *et al.* (1981).

As expected for a theory with a dimensional coupling constant, the use of Eq. (2.19) as a fundamental action leads to a nonrenormalizable perturbation expansion in G_F. This difficulty is resolved in the Weinberg-Salam gauge theory, in which in addition to the fermions, the fundamental action contains gauge and Higgs boson fields, and which has a renormalizable perturbation expansion in the gauge boson couplings g,g'. When the boson fields are integrated out, according to

$$e^{iS_{\text{eff}}[\{\text{fermions}\}]} = \int d\{\text{bosons}\} e^{iS_{\text{Weinberg-Salam}}[\{\text{fermions}\},\{\text{bosons}\}]} , \quad (2.21)$$

the effective action of Eq. (2.19) is obtained as a leading approximation, with the Fermi constant related to the electric charge e, the charged gauge boson mass M_W, and $\sin\theta_W$ by

$$\frac{1}{\sqrt{2}} G_F = \frac{g^2}{8M_W^2} , \qquad g = \frac{e}{\sin\theta_W} . \quad (2.22)$$

Let us now return to gravitation. The action of Eq. (1.1) contains dimensional couplings G^{-1} and Λ,

$$\begin{aligned} &\dim[G^{-1}] = \dim[\Lambda] = 2 , \\ &G^{-1/2} = m_{\text{Planck}} = 1.22\times 10^{19}\ \text{GeV} = l_{\text{Planck}}^{-1} , \\ &l_{\text{Planck}} = 1.62\times 10^{-33}\ \text{cm} , \\ &|\Lambda| \lesssim 10^{-57}\ \text{cm}^{-2} , \end{aligned} \quad (2.23)$$

and, as expected for the case when the couplings are not dimensionless, leads to a nonrenormalizable quantum field theory. The viewpoint of this article will be that the gravitational action is not a fundamental microscopic action, but rather is a long-wavelength effective action similar to the ones discussed above. The fundamental action will be assumed to be renormalizable, and conditions on it will be formulated which guarantee that the effective gravitational action is calculable in terms of parameters of the microscopic theory.

C. Renormalizability and the dimensional algorithm

In quantum field theory, one in general encounters divergences when evaluating radiative corrections. In renormalizable field theories, all divergences can be eliminated by making divergent rescalings, or renormalizations, of a finite number of parameters of the theory, which cannot be calculated from first principles but are replaced by measured values at the end of the calculation.

For example, in spin-$\frac{1}{2}$ quantum electrodynamics, working for simplicity to one-loop order, one introduces renormalization constants $Z_{1,2,3,m,e}$, renormalized fields A_μ^r, $F_{\mu\nu}^r$, ψ^r, and renormalized charge and mass parameters e,m by writing

$$\begin{aligned} &A_\mu = Z_3^{1/2} A_\mu^r, \quad F_{\mu\nu} = Z_3^{1/2} F_{\mu\nu}^r , \\ &\psi = Z_2^{1/2} \psi^r , \\ &e_0 = Z_e^{1/2} e, \quad m_0 = Z_m m ; \end{aligned} \quad (2.24a)$$

$$\begin{aligned} &F_{\mu\nu}F^{\mu\nu} = Z_3 F_{\mu\nu}^r F^{r\mu\nu} , \\ &\bar\psi\gamma^\mu\partial_\mu\psi = Z_2 \bar\psi^r \gamma^\mu \partial_\mu \psi^r , \\ &\bar\psi\gamma^\mu e_0 A_\mu \psi = Z_1 \bar\psi^r\gamma^\mu e A_\mu^r \psi^r = Z_2 Z_e^{1/2} Z_3^{1/2} \bar\psi^r \gamma^\mu e A_\mu^r \psi^r , \\ &\bar\psi m_0 \psi = Z_2 Z_m \bar\psi^r m \psi^r . \end{aligned} \quad (2.24b)$$

From Eq. (2.24b) and the Ward identity (which is derived from current conservation) one learns that

$$Z_e^{1/2} Z_3^{1/2} = \frac{Z_1}{Z_2} = 1 , \quad (2.25)$$

leaving as the independent renormalization constants Z_e, Z_m, and Z_2. Thus the renormalization procedure calls for the bare e_0, m_0, and ψ to be adjusted to absorb all divergences, leaving finite e, m, and ψ^r to be identified with the measured values. To understand why e, for example, cannot be calculated, let us recall that in one-loop order, the divergence in Z_e has the form

$$Z_e = 1 + \frac{\alpha_0}{3\pi}\log M^2 + O(\alpha_0^2), \quad \alpha_0 = \frac{e_0^2}{4\pi} , \quad (2.26)$$

with M a massive regulator. Under rescalings $M \to \xi M$ of the regulator mass, Z_e changes to

$$Z_e \to 1 + \frac{\alpha_0}{3\pi}\log M^2 + \frac{\alpha_0}{3\pi}\log\xi^2 + O(\alpha_0^2) , \quad (2.27)$$

and hence the finite part of Z_e is regulator-scheme dependent. As a result, the finite quantity e extracted from the divergent bare charge e_0 remains a free parameter of the renormalized theory. In general in a renormalizable field theory, we expect to find one free renormalized coupling or mass parameter for each bare coupling or mass appearing in the unrenormalized Lagrangian density.

Continuing for the moment to work to one-loop order, the renormalization procedure given in Eq. (2.24b) for the various dimension-four terms in the action density can be rewritten in a compact matrix notation,

$$[\Psi] = [Z][\Psi^r] ,$$

$$[\Psi] = \begin{bmatrix} F_{\mu\nu}F^{\mu\nu} \\ \bar\psi\gamma^\mu\partial_\mu\psi \\ \bar\psi\gamma^\mu e_0 A_\mu\psi \\ \bar\psi m_0\psi \end{bmatrix} , \quad [\Psi^r] = \begin{bmatrix} F_{\mu\nu}^r F^{r\mu\nu} \\ \bar\psi^r\gamma^\mu\partial_\mu\psi^r \\ \bar\psi^r\gamma^\mu e A_\mu^r\psi^r \\ \bar\psi^r m\psi^r \end{bmatrix} .$$

$$[Z] = \begin{bmatrix} Z_3 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & Z_1 & 0 \\ 0 & 0 & 0 & Z_2 Z_m \end{bmatrix} . \quad (2.28a)$$

Beyond one-loop order, the renormalization constants associated with the action density terms $F_{\mu\nu}F^{\mu\nu}, \ldots$ are no longer simply products of the renormalization constants

for the individual field, charge, and mass factors introduced in Eq. (2.24a), and the action density terms themselves will mix under renormalization. The appropriate generalization of Eq. (2.28a) then takes the form

$$[\Psi]=[Z][\Psi^r],$$

$$[\Psi^r]=\begin{bmatrix}(F_{\mu\nu}F^{\mu\nu})^r\\ (\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi)^r\\ (\bar{\psi}\gamma^{\mu}e_0A_{\mu}\psi)^r\\ (\bar{\psi}m_0\psi)^r\end{bmatrix}, \tag{2.28b}$$

with $[\Psi]$ as in Eq. (2.28a), with $(F_{\mu\nu}F^{\mu\nu})^r,\ldots$, the renormalized action density terms, and with $[Z]$ a nondiagonal renormalization matrix.

As we have seen from the above example, in the generic case multiplicative renormalization takes the more general form of matrix multiplicative renormalization. The set of operators which can mix under this renormalization process is characterized by the following rule.

The dimensional algorithm [see Weinberg (1957), Zimmermann (1970), and Brown (1980)]. A composite operator in quantum field theory is defined (up to a constant factor) as the product of any number of fields or field derivatives at the same space-time point. The dimensional algorithm states: (i) The most general basis set of composite operators which can mix under renormalization are the polynomials of the same canonical dimension, and of the same symmetry type (spatial and internal) formed from the bare fields, the bare masses, and $\partial/\partial x^{\mu}$. (ii) The Lagrangian density for a renormalizable field theory must contain a complete basis set (apart from total derivatives) of Lorentz- and internal symmetry-invariant composite operators of canonical dimension four.

Let us illustrate the dimensional algorithm in the flat space-time cases of scalar φ^4 theory, QED$\frac{1}{2}$, and QCD, and then use it to deduce additional Lagrangian counterterms which must be added to assure renormalizability when these theories are embedded in a curved background manifold.

1. Scalar φ^4 theory in flat space-time

Excluding total derivatives, the only dimension-four composites even under $\varphi\to-\varphi$ (the internal symmetry of the model) are

$$\partial_{\mu}\varphi\partial^{\mu}\varphi,\; m_0^2\varphi^2,\; \varphi^4, \tag{2.29a}$$

$$m_0^4. \tag{2.29b}$$

The operators of Eq. (2.29a) are just the ones appearing in the Lagrangian density of Eq. (2.4), while in flat space-time Eq. (2.29b) is an irrelevant constant which can be dropped.

2. QED $\frac{1}{2}$ and QCD in flat space-time

For QED $\frac{1}{2}$, the only dimension-four composites (excluding total derivatives) are

$$F_{\mu\nu}F^{\mu\nu},\; \bar{\psi}\gamma^{\mu}D_{\mu}\psi,\; m_0\bar{\psi}\psi, \tag{2.30a}$$

$$m_0^4, \tag{2.30b}$$

$$\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi,\; m_0^2A_{\mu}A^{\mu},\; A_{\mu}\partial^2A^{\mu},\; (\partial_{\mu}A^{\mu})^2. \tag{2.30c}$$

The operators of Eq. (2.30a) are just the ones appearing in the Lagrangian density of Eq. (2.10), while in flat space-time Eq. (2.30b) is an irrelevant constant. The operators of Eq. (2.30c) are Lorentz scalars, but are not invariant under the internal symmetry (or gauge) transformation

$$A_{\mu}\to A_{\mu}+\partial_{\mu}\Phi,$$

$$\psi\to e^{-ie_0\Phi}\psi, \tag{2.31}$$

and hence do not appear in the Lagrangian density. For QCD the classification of gauge-invariant Lorentz scalar operators constructed from the bare fields is analogous—one simply adds an internal symmetry index i, and changes the definition of the covariant derivative as in Eq. (2.12). A careful proof that $A_{\mu i}A^{\mu i}$ is not an internal symmetry invariant in the non-Abelian case, taking account of the complexities introduced by gauge-fixing and ghost terms, is given in Appendix A.

3. Additional Lagrangian density terms in a background curved space-time (Brown and Collins, 1980)

When spin-0, spin-$\frac{1}{2}$, or gauge spin-1 matter fields are quantized on a curved background manifold with metric $g_{\mu\nu}$, the action takes the form

$$S[\{\phi\},g_{\mu\nu}]=\int d^4x\sqrt{-g}\,\mathscr{L}[\{\phi\},g_{\mu\nu}], \tag{2.32}$$

with $d^4x\sqrt{-g}$ the invariant volume element, and with $\mathscr{L}$ a scalar with respect to general-coordinate transformations. According to the dimensional algorithm, $\mathscr{L}$ must contain all scalar dimension-four polynomials which can be formed from the bare fields (including now $g_{\mu\nu}$), the bare masses, and $\partial/\partial x^{\mu}$, and which are invariant under the internal symmetries of the matter fields. The terms which can thus appear in $\mathscr{L}$ are easily enumerated, and may be conveniently grouped into the following four classes: (i) The generally covariant transcriptions of the Lagrangian densities of Eqs. (2.4)–(2.13), obtained in the usual manner by replacing ordinary derivatives ∂_{μ} by covariant derivatives ∇_{μ} with respect to the background metric. (ii) The bare mass terms m_0^4 of Eqs. (2.29b) and (2.30b), which contribute to the cosmological constant on a curved manifold,[3] as well

[3]The terms m_0^4 and m_0^2R, which appear in $\mathscr{L}$ multiplied by independent renormalization constants, may be considered, respectively, as the bare cosmological constant Λ_0/G_0 and the bare order-R Lagrangian density R/G_0. Prior to the discussion of the cosmological constant in Sec. VI.C, we shall not introduce bare parameters Λ_0, G_0 when not required to do so by the presence of dimensional parameters in the microscopic matter action.

as corresponding regulator mass terms M^4 if massive regulators are employed. (iii) Terms of first degree in the Riemann curvature tensor,[4]

$$\mathscr{O}_2 R \ , \tag{2.33}$$

with $\mathscr{O}_2$ a general-coordinate—scalar and internal symmetry-invariant operator of canonical dimension two. The allowed forms for $\mathscr{O}_2$ are[3]

$$m_0^2,\ M^2,\ \varphi^2 \ , \tag{2.34}$$

since as shown in Appendix A, $A_{\mu i}A^{\mu i}$ is excluded by gauge invariance. The differential operator $\mathscr{O}_2=\nabla_\mu\nabla^\mu$ is omitted from the list because $\sqrt{-g}\,\nabla_\mu\nabla^\mu R$ is a total derivative, and $\mathscr{O}_2=\nabla_\mu A^\mu$, with A^μ an Abelian gauge potential, is omitted because it is not gauge invariant. Moreover, as is also shown in Appendix A, the operator φ^2 is excluded by supersymmetry invariance when φ is a spin-0 partner of a massless supermultiplet.[5] (iv) Terms of second degree in the Riemann curvature tensor,

$$\begin{aligned}&\mathscr{G}=R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma}-4R_{\alpha\beta}R^{\alpha\beta}+R^2 \ ,\\&\mathscr{H}=C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma},\ \ \mathscr{K}=R^2 \ ,\end{aligned} \tag{2.35}$$

with $\mathscr{G}$ the Gauss-Bonnet density and $C_{\mu\nu\lambda\sigma}$ the Weyl conformal tensor, which in four dimensions has the form

$$\begin{aligned}C_{\mu\nu\lambda\sigma}=&\ R_{\mu\nu\lambda\sigma}-\tfrac{1}{2}(g_{\mu\lambda}R_{\nu\sigma}-g_{\mu\sigma}R_{\nu\lambda}-g_{\nu\lambda}R_{\mu\sigma}+g_{\nu\sigma}R_{\mu\lambda})\\&+\tfrac{1}{6}R(g_{\mu\lambda}g_{\nu\sigma}-g_{\mu\sigma}g_{\nu\lambda}) \ .\end{aligned} \tag{2.36}$$

The results of this enumeration can be summarized in the following lemmas:

Lemma 1. For a general renormalizable matter field theory (spin-0+spin-$\frac{1}{2}$+ gauge spin-1 fields) in curved space-time, quantized in a manner which respects all gauge and supersymmetry internal symmetries, the Lagrangian density terms proportional to R are of the following types,

$$\begin{aligned}&m_0^2R,\ \ m_0=\text{a bare mass} \ ,\\&M^2R,\ \ M=\text{a massive regulator} \ ,\\&\varphi^2R,\ \ \varphi=\text{a spin}-0\text{ field not a member}\\&\qquad\qquad\text{of a massless supermultiplet} \ .\end{aligned} \tag{2.37}$$

Lemma 2. If there are no bare masses or massive regulators and if all spin-0 fields belong to massless supermultiplets, then there are no terms in $\mathscr{L}$ proportional to R—that is, terms (iii) above are absent. Moreover, when these conditions are satisfied, terms (ii) above are also absent, and the structure of $\mathscr{L}$ reduces to

$$\begin{aligned}&\mathscr{L}[\{\phi\},g_{\mu\nu}]=\mathscr{L}_{\rm matter}[\{\phi\},g_{\mu\nu}]+\mathscr{L}_{\rm grav}[g_{\mu\nu}] \ ,\\&\mathscr{L}_{\rm grav}=A_0\mathscr{G}+B_0\mathscr{H}+C_0\mathscr{K} \ ,\end{aligned} \tag{2.38}$$

with $\mathscr{L}_{\rm matter}[\{\phi\},g_{\mu\nu}]$ the generally covariant transcription of the flat space-time matter Lagrangian density $\mathscr{L}[\{\phi\}]$. The splitting of $\mathscr{L}$ into the "matter" and "gravitational" parts given in Eq. (2.38) is unique, since in the absence of dimensional constants $\mathscr{L}_{\rm matter}$ and $\mathscr{L}_{\rm grav}$ satisfy

$$\begin{aligned}&\mathscr{L}_{\rm matter}[\{0\},g_{\mu\nu}]=0 \ ,\\&\mathscr{L}_{\rm grav}[\eta_{\mu\nu}]=0 \ ,\ \ \mathscr{L}_{\rm matter}[\{\phi\},\eta_{\mu\nu}]=\mathscr{L}[\{\phi\}] \ .\end{aligned} \tag{2.39}$$

D. Conditions for calculability of the gravitational effective action

We are now ready to return to a discussion of the gravitational effective action induced by quantized matter fields on a curved background. Following Eq. (2.16), we define the gravitational effective action by

$$e^{iS_{\rm eff}[g_{\mu\nu}]}=\int d\{\phi\}\,e^{iS[\{\phi\},g_{\mu\nu}]} \ . \tag{2.40}$$

Since $S_{\rm eff}$ is a scalar under general-coordinate transformations, it may be represented as the integral over the manifold of a scalar density, which for slowly varying metrics can be formally developed in a series expansion in powers of $\partial_\lambda g_{\mu\nu}$,[6]

$$\begin{aligned}&S_{\rm eff}[g_{\mu\nu}]=\int d^4x\,\sqrt{-g}\,\mathscr{L}_{\rm eff}[g_{\mu\nu}] \ ,\\&\mathscr{L}_{\rm eff}[g_{\mu\nu}]=\mathscr{L}_{\rm eff}^{(0)}[g_{\mu\nu}]+\mathscr{L}_{\rm eff}^{(2)}[g_{\mu\nu}]+O[(\partial_\lambda g_{\mu\nu})^4] \ ,\\&\mathscr{L}_{\rm eff}^{(0)}[g_{\mu\nu}]=\frac{1}{16\pi G_{\rm ind}}(-2\Lambda_{\rm ind}),\ \ \mathscr{L}_{\rm eff}^{(2)}[g_{\mu\nu}]=\frac{1}{16\pi G_{\rm ind}}R \ .\end{aligned} \tag{2.41}$$

What are the conditions for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ to be uniquely calculable in terms of the renormalized parameters of the flat space-time matter theory? Clearly, if the fundamental action $S[\{\phi\},g_{\mu\nu}]$ contains terms proportional to R, then the finite renormalization ambiguities arising from these terms will produce an undetermined finite contribution to $G_{\rm ind}^{-1}$; in this case the induced gravitational con-

[4]No additional counterterms of first order in the curvature tensor can be formed by using the Ricci tensor $R_{\mu\nu}$, since these must have the form $\mathscr{O}_2^{\mu\nu}R_{\mu\nu}$, with $\mathscr{O}_2^{\mu\nu}$ a rank-two symmetric tensor of canonical dimension two. The only possibilities are $\mathscr{O}_2^{\mu\nu}=\nabla^\mu A^\nu+\nabla^\nu A^\mu$, which can be reduced to $\nabla_\mu A^\mu R$ by integration by parts and use of the Bianchi identity $\nabla^\mu R_{\mu\nu}=\frac{1}{2}\nabla_\nu R$, and $\mathscr{O}_2^{\mu\nu}=\mathscr{O}_2 g^{\mu\nu}$, which is equivalent to Eq. (2.33) of the text. Similarly, no additional counterterms can be formed by using the Weyl conformal tensor $C_{\mu\nu\lambda\sigma}$, and so the enumeration given in the text is complete.

[5]See Fayet and Ferrara (1977) for a discussion of supersymmetry field representations.

[6]In renormalizable theories, massless particle loops in general give rise to logarithms of $\partial_\lambda g_{\mu\nu}$ in the $(\partial_\lambda g_{\mu\nu})^4$ terms (that is, in the curvature-squared terms) of Eq. (2.41). For example, the existence of a conformal trace anomaly proportional to $\mathscr{H}$ indicates the presence of an effective action term proportional to $\int d^4x\,\sqrt{-g}\,\mathscr{H}\log\mathscr{H}$.

stant is renormalizable, but not calculable. On the other hand, if no terms proportional to R appear in $S[\{\phi\},g_{\mu\nu}]$, then $G_{\rm ind}^{-1}$ will be calculable, since there will now be no source of ambiguity proportional to R.[7] In this case the theory will yield a uniquely determined finite value for $G_{\rm ind}^{-1}$. So we have the following result:

Theorem [Adler (1980a)]. Under the conditions of lemma 2, a quantized matter theory in a curved background produces a calculable induced gravitational constant $G_{\rm ind}^{-1}$.

Consider next the induced cosmological constant $\Lambda_{\rm ind}$, which appears in the effective action in the dimension-four combination $\Lambda_{\rm ind}/G_{\rm ind}$. Ambiguities in $\Lambda_{\rm ind}$ can arise only from dimension-four terms in the flat space-time limit of $S[\{\phi\},g_{\mu\nu}]$ which are not determined by the renormalization conditions on the flat space-time matter theory. The decomposition of Eqs. (2.38)–(2.39) guarantees that no such additional dimension-four terms are present, and so we can conclude:

Theorem. Under the conditions of lemma 2, a quantized matter theory in a curved background produces a calculable induced cosmological constant $\Lambda_{\rm ind}$, and so the entire effective Einstein-Hilbert gravitational action is calculable.

The basic theorems just stated give *sufficient conditions* for the finiteness of the induced gravitational action. Of the three conditions in lemma 1, two—the absence of bare masses, and of scalar fields not in massless supermultiplets—are also necessary conditions. However, the exclusion of massive regulators is not necessary, and in Appendix A the analysis is generalized to the case where massive regulators are employed. As discussed in Sec. 4.4 of Fadde'ev and Slavnov (1980), massive regulators have useful formal properties, but they are awkward to use in explicit calculations. A superior method for diagram evaluations is the technique of dimensional regularization, which is discussed in Sec. III below. The subsequent sections of this review contain elaborations on the theorems of this section. For the theorems to have a nontrivial content, we must have a way of generating a nonzero scale for physical masses even when bare masses are zero (otherwise we get $G_{\rm ind}^{-1}=0$, which is calculable but trivial); this requires dynamical breaking of scale invariance, as discussed in detail in Sec. IV. In Sec. V, we derive explicit, formal expressions for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in terms of expectations of operators in the flat space-time matter vacuum. Finally, in Sec. VI, I extend the discussion to include the effects of quantization of the metric.

III. DIMENSIONAL REGULARIZATION

A. Survey

The regularization of quantum field theory without introducing massive regulators can be accomplished by analytic regularization methods, in which divergent integrals are defined by analytic continuation in a dimensionless parameter (for a review, see Leibbrandt, 1975). It will suffice to limit the discussion to regularization methods for flat space-time, since we will see below (in Sec. V) that after doing the curvature arithmetic needed to extract expressions for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$, we can explicitly take the flat space-time limit in the resulting formulas. The most widely used form of analytic regularization for flat space-time calculations is dimensional regularization, in which the dimension of the space-time manifold is continued from 4 to 2ω by the coordinate and momentum space replacements

$$\int d^4x \to \int d^{2\omega}x$$
$$\int d^4p \to \int d^{2\omega}p \; , \tag{3.1}$$

while keeping the formal structure of the action, in terms of fields and field derivatives, the same as in dimension four. After Wick rotation to 2ω-dimensional Euclidean space,[8] Feynman integrands in the continued theory are evaluated by using the following simple rules. The Kronecker delta δ^μ_ν obeys the usual composition law

$$\delta^\mu_\nu\delta^\nu_\sigma=\delta^\mu_\sigma \; , \tag{3.2a}$$

but its trace is modified to

$$\delta^\mu_\mu=2\omega \; . \tag{3.2b}$$

From Eq. (3.2) the symmetric average of momentum factors can be uniquely deduced, giving, for example,

$$\langle p_\mu p_\nu\rangle_{\substack{\text{symmetric}\\ \text{average}}}=\frac{p^2}{2\omega}\delta_{\mu\nu} \; . \tag{3.3}$$

The Dirac γ matrices continue to obey a Clifford algebra

$$\{\gamma_\mu,\gamma_\nu\}=2\delta_{\mu\nu}1 \; , \tag{3.4a}$$

and are trace normalized so that

$$\mathrm{Tr}(\gamma_\mu\gamma_\nu)=2^\omega\delta_{\mu\nu} \; , \quad \mathrm{Tr}(1)=2^\omega \; , \tag{3.4b}$$

permitting one to deduce unique values for all spinor loops not containing an odd number of factors γ_5.[9] Using Eqs. (3.2)–(3.4) and rotational covariance, all perturbation theory calculations can be reduced to multiple integrals of scalar-valued integrands over the momentum space of dimension 2ω.

The basic momentum space integral which appears is

$$\int_E d^{2\omega}p\, f(p) \; , \tag{3.5}$$

and is uniquely specified, up to an overall normalization, by the following three conditions given by Wilson (1973),

[7]This type of argument was first used in connection with the calculability of mass relations by Weinberg (1972).

[8]I will use the notation $\int_E$ to denote a Euclidean integral, and will consistently use a $++++$ metric convention in Euclidean space formulas.

[9]For recent discussions of the dimensional regularization treatment of γ_5, see Gottlieb and Donohue (1979) and Ovrut (1981).

linearity:

$$\int_E d^{2\omega}p[af_1(p)+bf_2(p)]=a\int_E d^{2\omega}p\, f_1(p) + b\int_E d^{2\omega}p\, f_2(p)\,,$$

translation invariance:

$$\int_E d^{2\omega}p\, f(p+q)=\int_E d^{2\omega}p\, f(p)\,,$$

scaling law:

$$\int_E d^{2\omega}p\, f(sp)=s^{-2\omega}\int_E d^{2\omega}p\, f(p)\,. \tag{3.6}$$

The normalization which is conventionally used is

$$\int_E d^{2\omega}p\, e^{-p^2}=\pi^\omega\,, \tag{3.7}$$

but is fixed only at $\omega=2$; Collins (1975) shows that ambiguities in normalization away from $\omega=2$ can always be absorbed into the ambiguities of the renormalization constants discussed in Sec. II.C. Hence dimensional regularization gives a well-defined procedure for regularizing the ultraviolet divergences of quantum field theory.[10]

Using the rules of Eqs. (3.5)–(3.7), we find, for example, that

$$\int_E d^{2\omega}p(p^2+m^2)^{-\alpha}=\pi^\omega\frac{\Gamma(\alpha-\omega)}{\Gamma(\alpha)}(m^2)^{\omega-\alpha}\,. \tag{3.8}$$

For $\omega-\alpha<0$, the integral on the left is convergent in the ultraviolet and yields the expression on the right, which is meromorphic (analytic apart from isolated poles) in ω and α. The integral can then be defined by analytic continuation for $\omega-\alpha>0$, except at points where it develops poles. For example, when $\alpha=1$ we have

$$\int_E d^{2\omega}p\frac{1}{p^2+m^2}=\pi^\omega\Gamma(1-\omega)(m^2)^{\omega-1} = \begin{cases} \dfrac{\pi}{1-\omega}+\text{finite}\,, & \text{near } \omega=1 \\ \left[-\dfrac{\pi^2}{2-\omega}+\text{finite}\right]m^2, & \text{near } \omega=2\,, \end{cases} \tag{3.9}$$

showing that the pole at $\omega=1$ is associated with the two-dimensional logarithmically divergent integral

$$\int_E\frac{d^2p}{p^2}\sim\frac{\pi}{1-\omega},\quad \omega\to1\,, \tag{3.10a}$$

while the pole at $\omega=2$ is associated with the four-dimensional logarithmically divergent integral

$$\int_E d^4p\frac{1}{(p^2)^2}\sim\frac{\pi^2}{2-\omega},\quad \omega\to2\,. \tag{3.10b}$$

The integral of Eq. (3.10b) would be represented by $\pi^2\log M^2$ using a conventional massive regulator, giving the useful correspondence

$$\log M^2 \leftrightarrow \frac{1}{2-\omega} \tag{3.11}$$

between the representation of a four-dimensional logarithmic divergence in the massive regulator and the dimensional regularization schemes. In N-loop order in dimensional regularization, one in general encounters higher powers of logarithmic divergences $1/(2-\omega),\ldots,1/(2-\omega)^N$ near $\omega=2$; these divergences must be cancelled against corresponding poles in the renormalization constants Z in order to extract finite physical amplitudes at dimension four.

B. Vanishing of quadratic divergences

The formally quadratically divergent (and m^2-independent) integral

$$\int_E\frac{d^4p}{p^2} \tag{3.12}$$

is assigned the value 0 by dimensional regularization, since the right-hand side of Eq. (3.9) is proportional to m^2 at $\omega=2$. A more precise statement of this fact is given by the following:

Lemma. The only evaluation of the ultraviolet divergent, infrared convergent massless integral

$$I^{\omega,\alpha}=\int_E d^{2\omega}p(p^2)^{-\alpha},\quad \omega-\alpha>0\,, \tag{3.13}$$

which is meromorphic in ω and α and which agrees with the $m\to0$ limit of Eq. (3.8), is $I^{\omega,\alpha}=0$. The proof follows immediately from the observations that: (i) when $\omega-\alpha>0$ is not a positive integer, the limit as $m\to0$ of Eq. (3.8) exists, and is 0; and (ii) the only meromorphic extension (to $\omega-\alpha=$positive integer) of 0 is 0.[11] Working from $I^{\omega,\alpha}=0$, we can now prove the vanishing of

$$I^{\omega,\alpha,\beta}=\int_E d^{2\omega}p(p^2)^{-\alpha}(\log p^2)^{-\beta},\quad \omega-\alpha>0 \tag{3.14}$$

by repeated differentiation of $I^{\omega,\alpha}$ with respect to α, and by repeated application of the Weyl transform (Erdélyi, 1954)

$$W^\beta I^{\omega,\alpha,\beta'}\equiv\frac{1}{\Gamma(\beta)}\int_\alpha^{i\infty}d\gamma(\gamma-\alpha)^{\beta-1}I^{\omega,\gamma,\beta'} = \int_E d^{2\omega}p[W^\beta(p^2)^{-\alpha}](\log p^2)^{-\beta'}\,. \tag{3.15}$$

Since

$$W^\beta(p^2)^{-\alpha}=\frac{1}{\Gamma(\beta)}(p^2)^{-\alpha}\int_0^{i\infty}d\delta\,\delta^{\beta-1}(p^2)^{-\delta} =(p^2)^{-\alpha}(\log p^2)^{-\beta},\quad 0<\beta<1\,, \tag{3.16}$$

we have

[10]A detailed axiomatization of the rules of dimensional regularization, along the lines sketched by Wilson (1973), has been given by Collins (unpublished).

[11]Any nonzero evaluation of $I^{\omega,\alpha}$ (such as the one given by Leibbrandt, 1975) is thus necessarily not a meromorphic function of ω. Such evaluations violate the basic philosophy of analytic regularization, which is essentially a calculus of meromorphic functions. The vanishing of $I^{\omega,\alpha}$ in dimensional regularization was first noted by 't Hooft and Veltman (1972), and is deduced as a theorem in the axiomatization of Collins (unpublished).

$$W^{\beta}I^{\omega,\alpha,\beta'}=I^{\omega,\alpha,\beta'+\beta},\quad 0<\beta<1\ ,$$

$$(-\partial/\partial\alpha)I^{\omega,\alpha,\beta'}=I^{\omega,\alpha,\beta'-1}\ , \tag{3.17}$$

and so by repeated operations any value of β can be reached, starting from $\beta=0$, where we have $I^{\omega,\alpha,0}=I^{\omega,\alpha}=0$. By continuing this procedure with respect to the index β we can generate powers of $\log\log p^2$, etc., giving finally:

Lemma. In dimensional regularization, for $\omega-\alpha>0$ we have

$$I^{\omega,\alpha,\beta,\gamma,\cdots}=\int_E d^{2\omega}p(p^2)^{-\alpha}(\log p^2)^{-\beta}(\log\log p^2)^{-\gamma}\cdots$$
$$=0\ . \tag{3.18}$$

In particular, the generalized quadratically divergent integral vanishes,

$$I^{2,1,\beta,\gamma,\cdots}=\int_E \frac{d^4p}{p^2}(\log p^2)^{-\beta}(\log\log p^2)^{-\gamma}\cdots=0\ . \tag{3.19}$$

This result shows that computing radiative corrections to the basic quadratically divergent integral of Eq. (3.12) always gives 0, independently of whether one proceeds order by order in perturbation theory, which gives only positive powers of $\log p^2$ (corresponding to $I^{2,1,-n}$), or whether one uses the renormalization group to sum powers of $\log p^2$ into running coupling constant factors (see Sec. IV.C below), giving the more general integral of Eq. (3.18).

C. An application: the stress tensor trace anomaly in gauge theories

As an application of dimensional regularization, let us derive, to one-loop order, the flat space-time stress tensor trace anomaly in QED $\frac{1}{2}$. In spinor quantum electrodynamics, the symmetrized stress energy tensor is given by

$$T_{\mu\nu}=\tfrac{1}{4}\eta_{\mu\nu}F_{\lambda\sigma}F^{\lambda\sigma}-F_{\lambda\mu}F^{\lambda}{}_{\nu}$$
$$+\frac{i}{4}[\bar\psi(\gamma_\nu D_\mu+\gamma_\mu D_\nu)\psi-\bar\psi(\overleftarrow{D}_\mu\gamma_\nu+\overleftarrow{D}_\nu\gamma_\mu)\psi]\ ,$$
$$D_\mu=\partial_\mu+ie_0A_\mu,\quad \overleftarrow{D}_\mu=\overleftarrow{\partial}_\mu-ie_0A_\mu\ . \tag{3.20}$$

Contracting with $\eta^{\mu\nu}$ and using Eq. (3.2b) and the spinor equation of motion

$$i\gamma^\mu D_\mu\psi=m_0\psi\ ,$$
$$-i\bar\psi\overleftarrow{D}_\mu\gamma^\mu=m_0\bar\psi\ , \tag{3.21}$$

we get

$$T^\mu_\mu=-2(2-\omega)\tfrac{1}{4}F_{\lambda\sigma}F^{\lambda\sigma}+\bar\psi m_0\psi\ . \tag{3.22}$$

Although the first term on the right-hand side of Eq. (3.22) is proportional to $2-\omega$, it cannot be dropped as $\omega\rightarrow2$ because the factor $F_{\lambda\sigma}F^{\lambda\sigma}$ contains a pole series in $(2-\omega)^{-1}$. To exhibit these poles explicitly, Eqs. (2.24)–(2.26) and Eq. (3.11) are used to write

$$F_{\lambda\sigma}F^{\lambda\sigma}=Z_e^{-1}F^r_{\lambda\sigma}F^{r\lambda\sigma}\ ,$$
$$Z_e=1+\frac{\alpha_0}{3\pi}\log M^2+O(\alpha_0^2)$$
$$\leftrightarrow 1+\frac{\alpha_0}{3\pi}\frac{1}{2-\omega}+O(\alpha_0^2)\ , \tag{3.23}$$
$$Z_e^{-1}=1-\frac{\alpha_0}{3\pi}\frac{1}{2-\omega}+\left[\frac{\alpha_0}{3\pi}\right]^2\frac{1}{(2-\omega)^2}+\cdots\ ,$$

where we have worked to one-loop order in the photon proper self-energy, and to iterated one-loop order in Z_e^{-1}. Substituting Eq. (3.23) into the first term of Eq. (3.22), we get (in the limit as $\omega\rightarrow2$)

$$-2(2-\omega)Z_e^{-1}=\frac{2\alpha_0}{3\pi}\left[1-\frac{\alpha_0}{3\pi}\frac{1}{2-\omega}+\left[\frac{\alpha_0}{3\pi}\right]^2\frac{1}{(2-\omega)^2}+\cdots\right]$$
$$=\frac{2\alpha_0}{3\pi}Z_e^{-1}=\frac{2\alpha}{3\pi},\quad \alpha=\frac{e^2}{4\pi}\ . \tag{3.24}$$

Hence to one-loop order the stress energy tensor trace is

$$T^\mu_\mu=\frac{2\alpha}{3\pi}\tfrac{1}{4}F^r_{\lambda\sigma}F^{r\lambda\sigma}+\bar\psi m_0\psi\ . \tag{3.25}$$

The first term on the right-hand side, found by Coleman and Jackiw (1971), Crewther (1972), and Chanowitz and Ellis (1972, 1973), would be lost if one naively used the equations of motion without attention to regularization, and is called the trace anomaly. The derivation given above can be generalized to all orders in perturbation theory (Adler, Collins, and Duncan, 1977; Nielsen, 1977) and yields

$$T^\mu_\mu=\frac{2\beta(e)}{e}\tfrac{1}{4}(F_{\lambda\sigma}F^{\lambda\sigma})^r+[1+\delta(e)](\bar\psi m_0\psi)^r\ , \tag{3.26}$$

with β and δ finite functions of e which are defined through the renormalization group, and with the splitting of T^μ_μ into the two terms on the right-hand side made unique by the specification of certain zero momentum-transfer matrix elements of the composite operators $(F_{\lambda\sigma}F^{\lambda\sigma})^r$ and $(\bar\psi m_0\psi)^r$. The analogous formula for QCD (obtained by Collins, Duncan, and Joglekar, 1977 and Nielsen, 1977) reads[12]

$$T^\mu_\mu=\frac{2\beta(g)}{g}\tfrac{1}{4}(F^i_{\lambda\sigma}F^{i\lambda\sigma})^r+[1+\delta(g)](\bar\psi m_0\psi)^r\ . \tag{3.27}$$

For a pure SU(n) gauge theory with no quarks, the second term on the right-hand side is absent, and the trace anomaly formula simplifies to

[12]I have dropped equation of motion terms, which both vanish at nonzero momentum transfer and have vanishing zero-momentum-transfer vacuum expectation values.

$$T^{\mu}_{\mu}=\frac{2\beta(g)}{g}\tfrac{1}{4}(F^{l}_{\lambda\sigma}F^{l\lambda\sigma})^{r}\,. \tag{3.28}$$

Equation (3.28) will play an important role in the analysis, given in Sec. V.D below, of $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in an SU(n) gauge theory.

IV. SYMMETRY BREAKDOWN

A. Models with elementary scalars

Spontaneous symmetry breaking plays a crucial role in constructing gauge theory models, since it permits generation of the gauge boson masses needed to get realistic low-energy effective actions, while preserving the ultraviolet cancellations which guarantee renormalizability. The simplest model exhibiting spontaneous symmetry breaking is the scalar φ^4 field theory of Eq. (2.4),

$$\begin{aligned}&\mathscr{L}=T-V\,,\\&T=\tfrac{1}{2}(\partial_0\varphi)^2\,,\\&V=\tfrac{1}{2}(\partial_i\varphi)^2+\tfrac{1}{2}m_0^2\varphi^2+\tfrac{1}{4}\lambda_0\varphi^4\,.\end{aligned} \tag{4.1}$$

For constant φ, the potential V reduces to

$$V(\varphi)=\tfrac{1}{2}m_0^2\varphi^2+\tfrac{1}{4}\lambda_0\varphi^4\,, \tag{4.2}$$

and has the behavior sketched in Fig. 1. In the conventional case $m_0^2>0$, the potential has a single stable minimum at $\varphi=0$, as shown in Fig. 1(a). However, when the sign of m_0^2 is reversed to $m_0^2<0$, the extremum at $\varphi=0$ becomes unstable, and V develops a pair of stable minima at

$$\begin{aligned}&\varphi=\pm\bar{\varphi}\,,\\&\bar{\varphi}=(-m_0^2/\lambda_0)^{1/2}\,,\end{aligned} \tag{4.3}$$

as shown in Fig. 1(b). Either the minimum at $\varphi=\bar{\varphi}$ or the minimum at $\varphi=-\bar{\varphi}$ can be used as the zeroth-order approximation in a perturbation expansion, by making a shift

$$\varphi=\pm\bar{\varphi}+\varphi' \tag{4.4}$$

and taking φ' as the new field variable. Mixing between the two configurations is not possible, because in the limit of an infinite space-time volume, they are separated by an infinite quantum-mechanical tunneling barrier. Thus the discrete $\varphi\leftrightarrow-\varphi$ symmetry of the Lagrangian is broken by the choice of one of the two classical minima as the quantum mechanical vacuum state.

The simplest field theory model in which a continuous symmetry is broken is obtained by making φ a complex scalar field

$$\varphi=\varphi_1+i\varphi_2\,, \tag{4.5}$$

with Lagrangian density

$$\mathscr{L}=\tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi^*-\tfrac{1}{2}m_0^2\varphi^*\varphi-\tfrac{1}{4}\lambda_0(\varphi^*\varphi)^2\,. \tag{4.6}$$

When $m_0^2<0$ the potential V has the behavior sketched in Fig. 2, and a suitable quantum-mechanical vacuum is obtained by making the shift

$$\varphi\to\bar{\varphi}+\varphi'\,, \tag{4.7a}$$

with $\bar{\varphi}$ any complex constant scalar satisfying

$$|\bar{\varphi}|^2=-m_0^2/\lambda_0\,. \tag{4.7b}$$

In this case a continuous symmetry is broken, and the excitation φ' which generates an infinitesimal rotation of $\bar{\varphi}$ is a zero-mass Goldstone mode. When the complex scalar field of Eqs. (4.5)–(4.6) is minimally coupled to a spin-1 gauge field, the zero-mass mode decouples from the physical degrees of freedom, and the spin-1 field be-

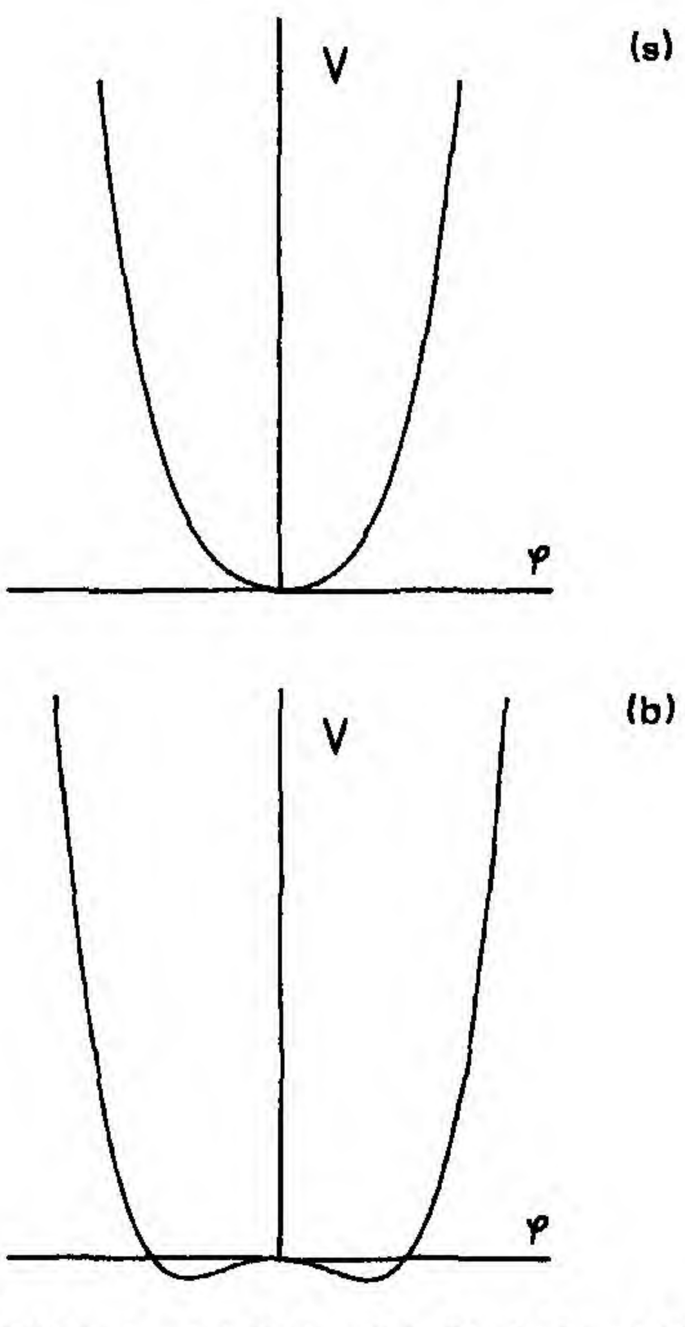


FIG. 1. (a) Potential V of Eq. (4.2), for $m_0^2>0$. (b) Potential V of Eq. (4.2), for $m_0^2<0$.

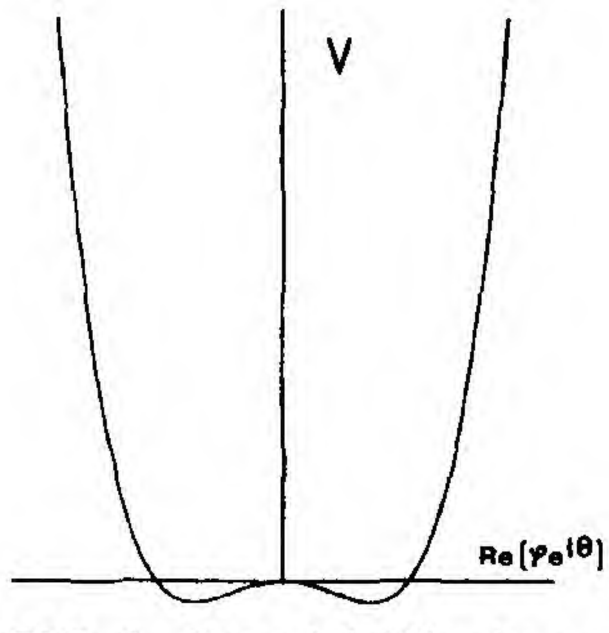


FIG. 2. Potential V in the complex scalar model, obtained by substituting $\varphi^2\to\varphi^*\varphi$ in Eq. (4.2). The curve shows an arbitrary section $\mathrm{Im}(\varphi e^{i\theta})=0$ of the potential surface, $0\le\theta<2\pi$, plotted vs $\mathrm{Re}(\varphi e^{i\theta})$.

comes massive. This is the so-called Higgs mechanism, which is used in the Weinberg-Salam model to generate intermediate vector boson masses. (For a detailed pedagogical review of these ideas and full references, see Bernstein, 1974).

The suggestion of linking spontaneous scale symmetry breaking with generation of the gravitational constant first appeared in the context of scalar meson models [see Fujii (1974), Englert, Truffin, and Gastmans (1976), Minkowski (1977), Chudnovsky (1978), Matsuki (1978), Smolin (1979), Zee (1979), Linde (1979, 1980) and Nieh (1982)]. The basic mechanism of the above-cited papers is to start from a Lagrangian density of the form

$$\mathscr{L}=\epsilon\varphi^2 R+T-V(\varphi^2)\ , \tag{4.8}$$

with V a symmetry-breaking potential as in Eq. (4.2). In the unstable symmetric phase $\bar\varphi=0$ there is no order-R term in $\mathscr{L}$, but in the stable broken-symmetry phase with $\bar\varphi^2=-m_0^2/\lambda_0$ an induced gravitational action is generated with

$$\frac{1}{16\pi G_{\text{ind}}}=\epsilon\bar\varphi^2\ . \tag{4.9}$$

In such models, since both scalar fields and dimensional parameters ($m_0\neq 0$) appear, the induced gravitational constant is not calculable[13]; ε is an additional curved space-time parameter of the theory which is not determined by the flat space-time renormalized parameters (Brown and Collins, 1980).

B. Dynamical symmetry breaking: the renormalization group in asymptotically free gauge theories

In order to get a calculable and nonvanishing induced gravitational constant, we must turn our attention to field theory models with dynamical scale-invariance breaking. Such theories, by definition, are formally scale invariant at the classical Lagrangian or tree-approximation level, but exhibit spontaneous scale-invariance breaking as a result of quantum corrections in one- or higher-loop order. There are two reasonably well understood mechanisms by which dynamical scale-invariance breaking can occur. The first, which will be discussed in this section, is through the renormalization process itself, in infrared-singular theories such as unbroken non-Abelian gauge theories. The second, which will be described in Sec. IV.C below, is through the generation of a mass gap and a fermion pair condensate in relativistic versions of the Bardeen-Cooper-Schrieffer (BCS, 1957) theory of superconductivity. The two mechanisms are not really disjoint, and both are believed to be operative in non-Abelian gauge theories. This fact and some further gauge theory-superconductor analogies are discussed briefly in Sec. IV.D. The material which follows has been organized so that the reader who wishes to proceed most directly to the gravitational applications of Secs. V and VI can do so after reading Sec. IV.B alone.

The most important class of field theory models exhibiting dynamical spontaneous scale-invariance breaking are asymptotically free gauge theories [see 't Hooft (unpublished), Gross and Wilczek (1973), and Politzer (1973)]. Consider an SU(n) non-Abelian gauge field coupled to N_f massless fermions in the fundamental representation, as is described, for example, by $\mathscr{L}_{\text{QCD}}$ of Eq. (2.12) with $m_0=0$ and with ψ replicated N_f times. In tree approximation this theory contains no dimensional parameters, and so scale invariance is unbroken; moreover, since there are no scalar fields, all of the conditions of the theorems of Sec. II are satisfied. Let us now consider the effect of quantum corrections to the tree-approximation theory. When radiative corrections are included, the coupling constant g appears in calculations through the running coupling constant

$$g^2(-q^2)=\frac{g^2(\mu^2)}{1+\frac{1}{2}b_0 g^2(\mu^2)\log(-q^2/\mu^2)+\cdots}\ , \tag{4.10}$$

with q^2 the four-momentum squared, μ^2 an arbitrary subtraction point, and $g^2(\mu^2)$ the value of the coupling constant at $-q^2=\mu^2$. The appearance of the subtraction mass μ^2 is necessitated by the fact that radiative corrections to massless gauge theories are highly infrared divergent, making it impossible to introduce a renormalized coupling parameter by specifying the value of g^2 at $q^2=0$, as is done in the more familiar case of quantum electrodynamics. The parameter b_0 is determined by one-loop radiative corrections to be

$$b_0=\frac{1}{8\pi^2}\left[\frac{11}{3}n-\frac{2}{3}N_f\right]\ , \tag{4.11}$$

and is positive, provided that N_f is not too large. When b_0 is positive, Eq. (4.10) shows that the running coupling vanishes at large four-momentum squared, and the theory in this case is said to be asymptotically free.

Let us examine the structure of Eq. (4.10) in the approximation in which only one-loop radiative corrections are retained, while the higher-loop contributions to the running coupling constant, denoted by ..., are neglected. (As discussed in Appendix B.1, there is a well-defined sense in which a one-loop analysis is exact.) Evaluating Eq. (4.10) at $-q^2=\mu_1^2$, we get

$$\frac{1}{g^2(\mu_1^2)}-\frac{1}{g^2(\mu^2)}=\frac{1}{2}b_0\log\left[\frac{\mu_1^2}{\mu^2}\right]$$

$$\Longrightarrow \frac{1}{g^2(\mu_1^2)}-\frac{1}{2}b_0\log\mu_1^2$$

$$=\frac{1}{g^2(\mu^2)}-\frac{1}{2}b_0\log\mu^2\ , \tag{4.12}$$

showing that the scale mass $\mathscr{M}(g(\mu),\mu)$ defined by

[13]Strictly speaking, to get a renormalizable model an additional term $\delta m_0^2 R$ must be included in $\mathscr{L}$; the spontaneous symmetry breaking then generates a change in the constant factor multiplying R from δm_0^2 to $\delta m_0^2+\epsilon\bar\varphi^2$.

$$\mathcal{M}(g(\mu),\mu)=\mu e^{-1/[b_0 g^2(\mu^2)]} \tag{4.13}$$

is subtraction-point independent. In technical terminology, the scale mass $\mathcal{M}(g(\mu),\mu)$ is said to be renormalization group[14] invariant to one-loop order, since it is left unchanged to this order by transformations of the renormalization point μ^2 and the renormalized coupling constant $g^2(\mu^2)$. When radiative corrections to all orders are kept, Eq. (4.13) generalizes to (Gross and Neveu, 1974; Lane, 1974a)

$$\mathcal{M}(g(\mu),\mu)=\mu e^{-\int^{g(\mu)} dg'/\beta(g')}, \tag{4.14}$$

with

$$\beta(g)=-\tfrac{1}{2}b_0 g^3+O(g^5) \tag{4.15}$$

the function appearing in the trace anomaly formula of Eq. (3.27), and again $\mathcal{M}(g,\mu)$ is said to be renormalization group invariant. An alternative, and frequently used, way of specifying that $\mathcal{M}$ has the functional form of Eq. (4.14) is obtained by requiring that $\mathcal{M}$ satisfy the Callan (1970)-Symanzik (1970) differential equation

$$\left[\mu\frac{\partial}{\partial\mu}+\beta(g)\frac{\partial}{\partial g}\right]\mathcal{M}(g,\mu)=0\,. \tag{4.16}$$

Let us now apply the above analysis to determine the structure of physically observable parameters, such as effective action parameters. Since observables must be subtraction-point independent, they can depend on μ only through the scale mass $\mathcal{M}(g,\mu)$, and so we get the following important result:

Theorem [Gross and Neveu (1974)]. Any physical parameter $P(g,\mu)$ which has canonical dimension d_P in the accounting of Sec. II.A must be equal to $[\mathcal{M}(g,\mu)]^{d_P}$ up to a calculable number,

$$P(g,\mu)=\text{calculable number}\times[\mathcal{M}(g,\mu)]^{d_P}\,. \tag{4.17}$$

Equivalently, $P(g,\mu)$ must satisfy the homogeneous renormalization group equation

$$\left[\mu\frac{\partial}{\partial\mu}+\beta(g)\frac{\partial}{\partial g}\right]P(g,\mu)=0\,, \tag{4.18}$$

which for a quantity of canonical dimension d_P implies Eq. (4.17).

According to this theorem, it is the dimensional scale mass $\mathcal{M}$, rather than the dimensionless (but subtraction-point dependent) renormalized coupling $g^2(\mu^2)$, which in asymptotically free gauge theories plays a role analogous to that played by the renormalized fine-structure constant α in quantum electrodynamics. In other words, the renormalization process has replaced a one-parameter family of unrenormalized theories, characterized by their values of the dimensionless unrenormalized gauge coupling g_0, by a one-parameter family of renormalized theories, characterized by their values of the dimension-one scale mass $\mathcal{M}(g,\mu)$. This change in dimensionality of the effective parameter, when radiative corrections are included, clearly implies that there has been a dynamical breaking of scale invariance. The general phenomenon is called dimensional transmutation, after Coleman and Weinberg (1973), who discovered similar behavior in massless QED 0 (a theory which, like the massless non-Abelian gauge theory, is highly infrared divergent.)

[14]For a pedagogical discussion of the renormalization group structure of non-Abelian gauge theories, see Stevenson (1981).

C. Dynamical symmetry breaking: relativistic generalizations of the superconductor gap equation

Historically, the earliest suggestion that dynamical symmetry breaking plays an important role in particle physics was contained in the classic paper of Nambu and Jona-Lasinio (1961), who proposed a model for nucleon mass generation[15] based on an analogy with the BCS theory of superconductivity.[16] The Nambu—Jona-Lasinio model starts from a Lagrangian containing massless, interacting fermions, and then sets up a self-consistent equation for the dynamically generated fermion mass in analogy with the "gap equation" of superconductivity. In this section, I give a very schematic account of the basic approximation method used in the BCS and Nambu—Jona-Lasinio models, and show that it gives a dynamical version of the tree-approximation model for symmetry breaking described in Sec. IV.A.

Let us consider a fermion with bare propagator G_0^{-1}, proper self-energy part Σ, and full propagator G^{-1}, related to one another as usual by

$$G^{-1}=G_0^{-1}-\Sigma\,. \tag{4.19}$$

Assuming the fermions interact through a potential V, a simple self-consistent approximation for the proper self-energy is obtained by truncating the Dyson equation for Σ to include only the lowest-order skeleton diagram illustrated in Fig. 3. This gives

$$\begin{aligned}\Sigma&=\int VG\\&=\int V(G_0^{-1}+\Sigma)[G_0^{-2}-\Sigma^2]^{-1}\,,\end{aligned} \tag{4.20}$$

where $\int$ indicates symbolically a summation or integration over intermediate state (closed loop) variables. In models with dynamical symmetry breaking, the unbroken symmetry of the classical Lagrangian can be shown to

[15]A very important aspect of the Nambu—Jona-Lasinio model, which is not dealt with in this review, is the generation of the pion as a zero-mass bound state. There has been recent interest in analogs of this phenomenon in which the Higgs scalars or pseudoscalars in unified theories are dynamically generated composites of more fundamental fields; see Englert and Brout (1964); Jackiw and Johnson (1973); Cornwall and Norton (1973); Weinberg (1976); and Susskind (1979).

[16]For texts on the BCS theory, see Schrieffer (1964) and Fetter and Walecka (1971). The Ginzburg-Landau phenomenological theory is also described in these books.

$\Sigma =$ 

FIG. 3. Truncated Dyson equation for the self-energy part. The dashed line and dots denote the potential V in the BCS case, or the photon propagator and emission and absorption vertices in the JBW model case. The heavy line denotes a full electron propagator $G=(G_0^{-1}-\Sigma)^{-1}$, giving a nonlinear integral equation (the gap equation) for Σ.

imply that

$$\int VG_0^{-1}[G_0^{-2}-\Sigma^2]^{-1}=0 \,. \tag{4.21}$$

Substituting Eq. (4.21) into Eq. (4.20) then gives the general form of the "gap equation" for Σ,

$$\Sigma=\int V\Sigma[G_0^{-2}-\Sigma^2]^{-1} \,. \tag{4.22}$$

Equation (4.22) always has a trivial solution $\Sigma=0$, analogous to the trivial root $\varphi=0$ of the equation

$$0=V'(\varphi)=\varphi(m_0^2+\lambda_0\varphi^2) \,, \tag{4.23}$$

which governs the vacuum structure of the scalar meson model discussed in Sec. IV.A. However, when V has the (attractive) sign for which dynamical symmetry breaking occurs, there is also a nontrivial solution to Eq. (4.22), corresponding symbolically to the root of

$$1=\int V[G_0^{-2}-\Sigma^2]^{-1} \,, \tag{4.24}$$

and analogous to the symmetry-breaking roots $\varphi=\pm\bar{\varphi}$ of Eq. (4.23).

To solve Eq. (4.24) explicitly in the case of the BCS model, we make substitutions appropriate to the nonrelativistic kinematics of the superconductor problem [see Schrieffer (1964)],

$$\begin{aligned} &\int = i\int \frac{dk_0}{\pi}\int d^3k \,, \\ &G_0^{-2}=k_0^2-(k^2-k_F^2)^2+i\varepsilon \,, \\ &\Sigma^2=\Delta^2 \,, \end{aligned} \tag{4.25}$$

where k_F is the Fermi momentum, and we carry out the k_0 integration. Equation (4.24) then yields an algebraic equation for the energy gap characterizing the low-lying electronic excitations in a superconductor,

$$1=V\int_{|k^2-k_F^2|=0}^{|k^2-k_F^2|=\omega_D} d^3k\frac{1}{[(k^2-k_F^2)^2+\Delta^2]^{1/2}} \,, \tag{4.26}$$

with ω_D the Debye frequency, which serves as an effective ultraviolet cutoff in the BCS model. Because phase space in the neighborhood of the Fermi momentum is effectively one dimensional,

$$d^3k \approx 4\pi k_F^2 dk \,, \tag{4.27}$$

Eq. (4.26) is logarithmically divergent at the lower limit when $\Delta=0$, and for small Δ can be approximated by

$$1=NV\int_{c\Delta}^{\omega_D}\frac{d\omega}{\omega}=NV\log\left[\frac{\omega_D}{c\Delta}\right] \,, \tag{4.28}$$

with N the density of states at the Fermi surface and c a numerical factor of order unity. Solving Eq. (4.28) for Δ gives

$$\Delta=\frac{1}{c}\omega_D\exp\left[-\frac{1}{NV}\right] \,, \tag{4.29}$$

showing that the energy gap has a nonperturbative dependence on the interaction strength V, with an essential singularity at $V=0$. The detailed analysis of the BCS model shows that the energy gap Δ is proportional to the ground-state expectation value of a product of creation (or annihilation) operators for two electrons, with opposite momenta lying near the Fermi surface and opposite spins,

$$\Delta \propto \left\langle \psi_{\mathbf{k}\uparrow}^{\dagger}\psi_{-\mathbf{k}\downarrow}^{\dagger}\right\rangle_0 \,, \quad |\mathbf{k}| \sim k_F \,. \tag{4.30}$$

Thus, the presence of a nonvanishing energy gap in a superconductor implies the existence of a ground-state condensate of correlated electron pairs.

An analogous reduction of Eq. (4.22) (now using relativistic kinematics) can be carried out for the Nambu–Jona-Lasinio model and for its more recent gauge-theoretic extensions, in which the nonrenormalizable local four-fermion interaction used by Nambu and Jona-Lasinio is replaced by a renormalizable interaction mediated by vector meson exchange. [See Johnson, Baker, and Willey (1964), Jackiw and Johnson (1973), Cornwall and Norton (1973), and Lane (1974b).] For definiteness, let us consider the case of the Johnson-Baker-Willey (JBW, 1964) model for fermion mass generation in Abelian electrodynamics. These authors consider zero-bare mass spinor electrodynamics [that is, $\mathscr{L}_{\mathrm{QED}\,1/2}$ of Eq. (2.10), with $m_0=0$] in the approximation in which all photon self-energy parts are neglected. The dashed line in Fig. 3 then represents a bare photon propagator; thus to leading order of perturbation theory for the vertex parts, the analog of Eq. (4.22) is

$$\begin{aligned} \Sigma(p) \sim i\alpha_0\int d^4k\frac{1}{k^2}\Sigma(p-k) \\ \times[(p-k)^2-\Sigma(p-k)^2]^{-1} \,, \end{aligned} \tag{4.31}$$

where $\sim$ indicates that numerical constants of order unity have been omitted. In addition to the trivial solution $\Sigma=0$, Eq. (4.31) has a nonperturbative solution in which Σ has the asymptotic behavior

$$\Sigma(p) \sim m\left[\frac{m^2}{-p^2}\right]^{\delta}, \quad \delta\sim\alpha_0 \,. \tag{4.32}$$

Equation (4.32) gives self-consistency because

$$\begin{aligned} &i\int d^4k\frac{1}{k^2}\frac{m}{(p-k)^2}\left[\frac{m^2}{-(p-k)^2}\right]^{\delta} \\ &\qquad \sim\frac{1}{\delta}m\left[\frac{m^2}{-p^2}\right]^{\delta} \sim \frac{1}{\alpha_0}\Sigma(p) \,, \end{aligned} \tag{4.33}$$

which follows from angular averaging and the elementary integral

$$\int_A^\infty \frac{dB}{B} B^{-\delta} = \frac{A^{-\delta}}{\delta} . \tag{4.34}$$

The parameter m in Eq. (4.32) is an arbitrary integration constant introduced by the boundary condition

$$\Sigma(p^2 = -m^2) = m , \tag{4.35}$$

and clearly corresponds to an electron physical mass. We see that as a result of dynamical symmetry breaking a mass scale has appeared in the solution to Eq. (4.31), even though no mass scale appears in the integral equation itself or in the fundamental Lagrangian from which it was derived. The vanishing of m_0 is mirrored in the fact that $\Sigma(p)$ has a softer ultraviolet behavior

$$\Sigma(p) \underset{p^2 \to \infty}{\longrightarrow} 0 \tag{4.36}$$

than would be found if a mass scale were introduced kinematically into the Lagrangian. Such ultraviolet softness (seen also in the discussion of asymptotically free gauge theories in Sec. IV.B above) is a very general feature of field theory models where the mass scale is introduced through dynamical scale-invariance breaking. The detailed analysis of the JBW and other Nambu–Jona-Lasinio type models shows that, associated with the generation of a nonvanishing fermion physical mass, the ground state contains a fermionic condensate, this time involving a nonvanishing fermion-antifermion expectation value of the form $\langle \bar{\psi}\psi \rangle_0$.

D. Gauge theory-superconductor analogies

Comparing Eq. (4.13) with Eq. (4.29), we see that there is a close similarity between the nonperturbative structure of the gauge theory one-loop scale mass $\mathcal{M}(g,\mu)$ and that of the superconductor energy gap Δ. As was noted in connection with Eqs. (4.26)–(4.28) above, the $e^{-1/NV}$ form in the superconductor case arises from the effectively one-dimensional phase space near the Fermi surface, which produces a logarithmically divergent one-loop perturbation theory contribution

$$\int \frac{d^3k}{k^2 - k_F^2} \sim \int_{k_F}^{k_{\max}} \frac{dk}{k - k_F} . \tag{4.37}$$

Similarly, the e^{-1/g^2} form in the gauge theory case arises from the logarithmic divergence of the one-loop contribution to $g^2(-q^2)^{-1}$ at $q^2 = 0$, which in turn comes from the nonvanishing and effectively one-dimensional phase space for a massless particle to decay into two massless particles, as expressed in the identity

$$\begin{aligned} &|\mathbf{k}_1|\,|\mathbf{k}_2|\,\delta^3(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)\delta(|\mathbf{k}|-|\mathbf{k}_1|-|\mathbf{k}_2|) \\ &\quad = 2\pi \int_0^1 dx\, \delta^3(\mathbf{k}_2 - \mathbf{k}x)\delta^3[\mathbf{k}_1 - \mathbf{k}(1-x)] . \end{aligned} \tag{4.38}$$

To see the effect of Eq. (4.38), let us recall that the S-wave phase space for a pair of particles of mass m, at center of mass energy $\sqrt{s}$, is

$$\rho(s) = \left[\frac{s - 4m^2}{s} \right]^{1/2} , \tag{4.39}$$

and vanishes at threshold for $m > 0$. However, when $m = 0$, Eq. (4.39) reduces to $\rho(s) = 1$, which is nonvanishing at threshold as required by Eq. (4.38). Consequently, the one-loop perturbation-theory integral

$$\int_0^{s_{\max}} \frac{ds' \rho(s')}{s' - q^2} \tag{4.40}$$

is logarithmically divergent at $q^2 = 0$.

As suggested by this phase-space analysis, and as discussed in more detail by Gross and Neveu (1974) and Lane (1974a, 1974b), the renormalization group mechanism for dynamical symmetry breaking on the one hand, and the superconductor gap equation mechanism on the other, are really two complementary aspects of the dynamical symmetry breaking which occurs in non-Abelian gauge theories. The gauge theory-superconductor analogy can be carried considerably further. Just as a superconductor contains an electron pair condensate proportional to the energy gap Δ, quantum chromodynamics contains a fermionic condensate $\langle \bar{\psi}\psi \rangle_0$ proportional to the third power $\mathcal{M}^3$ of the gauge theory scale mass $\mathcal{M}$, and very likely[17] contains a gluonic condensate $\langle F^l_{\lambda\sigma} F^{l\lambda\sigma} \rangle_0$ proportional to $\mathcal{M}^4$. When a superconductor and its energy gap are perturbed by a weakly varying electromagnetic field, the resulting dynamics is described by the induced effective action of the Ginzburg-Landau theory.[16] Correspondingly, when a non-Abelian gauge theory and its scale mass are perturbed by a weakly varying metric, the resulting dynamics, as we will see in detail below, is described by an induced effective action of the Einstein-Hilbert form.[18]

V. INDUCED GRAVITATIONAL AND COSMOLOGICAL CONSTANTS FOR MATTER THEORIES ON A BACKGROUND MANIFOLD

A. Path-integral derivation of formulas for G_{ind}^{-1} and Λ_{ind}

From the viewpoint of the theorem of Gross and Neveu discussed in Sec. IV.B, the induced gravitational

---

[17]For discussions of gluon pairing see Batalin, Matinyan, and Savvidi (1977); Savvidy (1977); Pagels and Tomboulis (1978); Vainstein, Zakharov, and Shifman (1978); Ambjörn and Olesen (1980); Fukuda and Kazama (1980); Kazama (1980); and Milton (1981). See also Sec. V.D below.

[18]The superconductor phase space analogy is discussed briefly in the "photon pairing" paper of Adler *et al.* (1976). One conclusion of their paper, that photon ladders cannot generate a graviton in flat space-time, is a special case of a recent general theorem of Witten and Weinberg (1980). The remainder of their paper and a subsequent paper of Adler (1976) attempted, unsuccessfully, to generate a gap equation as a curvature effect in a model which has no gap equation in the absence of curvature.

constant $G_{\rm ind}^{-1}$ and cosmological constant $\Lambda_{\rm ind}$ of a gauge field theory are simply physical parameters of canonical dimension two, defined through the response of the gauge field system to local perturbations in the space-time metric. This suggests that it should be possible to take formal derivatives with respect to deviations of the metric $g_{\mu\nu}$ from the Minkowski metric $\eta_{\mu\nu}$, thereby extracting expressions for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in terms of flat space-time vacuum expectation values. Such an analysis will be carried out in this section, using the metric and curvature conventions of Misner *et al.* (1970).

The starting point of the derivation is the basic definition of the gravitational effective action given in Eq. (2.40) above,

$$e^{iS_{\rm eff}[g_{\mu\nu}]}=\int d\{\phi\}e^{iS[\{\phi\},g_{\mu\nu}]}, \tag{5.1}$$

with

$$S_{\rm eff}[g_{\mu\nu}]=\int d^4x\sqrt{-g}\left[\frac{1}{16\pi G_{\rm ind}}(R-2\Lambda_{\rm ind})+O[(\partial_\lambda g_{\mu\nu})^4]\right], \tag{5.2a}$$

$$S[\{\phi\},g_{\mu\nu}]=\int d^4x\,\overline{\mathscr{L}}[\{\phi\},g_{\mu\nu}],$$

$$\overline{\mathscr{L}}[\{\phi\},g_{\mu\nu}]\equiv\sqrt{-g}\,\mathscr{L}[\{\phi\},g_{\mu\nu}]. \tag{5.2b}$$

I will assume that the microscopic action density $\overline{\mathscr{L}}$ is a function of the metric and its first and second derivatives,

$$\overline{\mathscr{L}}[\{\phi\},g_{\mu\nu}]=\mathscr{L}(\{\phi\},g_{\mu\nu},\partial_\lambda g_{\mu\nu},\partial_\lambda\partial_\sigma g_{\mu\nu}), \tag{5.3}$$

making the derivation general enough to encompass the case, discussed in Sec. VI below, where the metric itself (and not just the matter fields $\{\phi\}$) is path integral quantized. To proceed, let us calculate the conformal variation of Eq. (5.1) around a general background metric. This is done by acting on the left- and right-hand sides with the differential operator $2g_{\mu\nu}(y)\delta/\delta g_{\mu\nu}(y)$, where y is an arbitrary space-time point which will shortly be chosen as the origin, and then dividing by $i\exp(iS_{\rm eff})$. Inserting the expansion of Eq. (5.2a) in the left-hand side gives

$$2g_{\mu\nu}(y)\frac{\delta}{\delta g_{\mu\nu}(y)}\int d^4x\sqrt{-g}\left[\frac{1}{16\pi G_{\rm ind}}(R-2\Lambda_{\rm ind})+O[(\partial_\lambda g_{\mu\nu})^4]\right]=\frac{\int d\{\phi\}\,e^{iS[\{\phi\},g_{\mu\nu}]}2g_{\mu\nu}(y)\frac{\delta}{\delta g_{\mu\nu}(y)}\int d^4x\,\overline{\mathscr{L}}}{\int d\{\phi\}\,e^{iS[\{\phi\},g_{\mu\nu}]}}, \tag{5.4a}$$

where the quantities $g_{\mu\nu}$, $\overline{\mathscr{L}}$, R inside the x-integral are values at space-time point x, and where the functional integral $\int d\{\phi\}$ is still an integration over the values of the matter fields at all space-time points,

$$\int d\{\phi\}=\prod_z\int d\{\phi(z)\}. \tag{5.4b}$$

Equation (5.4a) can be evaluated using standard formulas for the first variations (with $T^{\mu\nu}$, as before, the renormalized matter stress-energy tensor),

$$\begin{aligned}\delta\sqrt{-g}&=\tfrac{1}{2}\sqrt{-g}\,g^{\mu\nu}\delta g_{\mu\nu},\\ \delta(\sqrt{-g}R)&=-\sqrt{-g}\,(R^{\mu\nu}-\tfrac{1}{2}g^{\mu\nu}R)\delta g_{\mu\nu}\\ &\quad+\text{total derivatives},\end{aligned} \tag{5.5}$$

$$\begin{aligned}\delta\overline{\mathscr{L}}&=\tfrac{1}{2}\overline{T}^{\mu\nu}\delta g_{\mu\nu},\\ \overline{T}^{\mu\nu}&\equiv\sqrt{-g}\,T^{\mu\nu}\\ &=2\left[\frac{\partial\overline{\mathscr{L}}}{\partial g_{\mu\nu}}-\partial_\lambda\frac{\partial\overline{\mathscr{L}}}{\partial(\partial_\lambda g_{\mu\nu})}+\partial_\lambda\partial_\sigma\frac{\partial\overline{\mathscr{L}}}{\partial(\partial_\lambda\partial_\sigma g_{\mu\nu})}\right].\end{aligned} \tag{5.6}$$

Substituting these, and defining the point y to be the origin 0 in order to simplify the subsequent formulas, we get

$$\frac{1}{8\pi G_{\rm ind}}[R(0)-4\Lambda_{\rm ind}]+O[(\partial_\lambda g_{\mu\nu})^4]=\frac{\int d\{\phi\}e^{iS[\{\phi\},g_{\mu\nu}]}\overline{T}[g_{\mu\nu},0]}{\int d\{\phi\}e^{iS[\{\phi\},g_{\mu\nu}]}}, \tag{5.7}$$

with $\overline{T}[g_{\mu\nu},x]$ the stress-energy tensor trace functional defined by

$$\begin{aligned}\overline{T}[g_{\mu\nu},x]&=\sqrt{-g}\,T_\mu^\mu\\ &=2g_{\mu\nu}\left[\frac{\partial\overline{\mathscr{L}}}{\partial g_{\mu\nu}}-\partial_\lambda\frac{\partial\overline{\mathscr{L}}}{\partial(\partial_\lambda g_{\mu\nu})}+\partial_\lambda\partial_\sigma\frac{\partial\overline{\mathscr{L}}}{\partial(\partial_\lambda\partial_\sigma g_{\mu\nu})}\right].\end{aligned} \tag{5.8a}$$

Taking the flat space-time limit of Eq. (5.7) and introducing the abbreviated notation

$$T(x)\equiv\overline{T}[\eta_{\mu\nu},x]=T_\mu^\mu\,|_{g_{\mu\nu}=\eta_{\mu\nu}}, \tag{5.8b}$$

we obtain a formula for the induced cosmological term,

$$-\frac{1}{2\pi}\frac{\Lambda_{\rm ind}}{G_{\rm ind}}=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(0)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}. \tag{5.9}$$

In order to extract a formula for the induced gravitational constant, we must take a further metric variation of Eq. (5.7). Since the left-hand side of Eq. (5.7) has no tensor structure, it suffices to specialize[19] to a metric which around $x=0$ has the conformally flat, constant-curvature form

[19]For a derivation which does not make this specialization, but instead proceeds from the general Riemann normal expansion $g_{\mu\nu}=\eta_{\mu\nu}-(\frac{1}{3})R_{\mu\alpha\nu\beta}x^\alpha x^\beta+\cdots$, see Adler (1980c). See also Brown and Zee (1982).

$$g_{\mu\nu}(x)=\eta_{\mu\nu}[1-\tfrac{1}{24}R(0)x^2+O(\nabla R,R^2)]$$
$$=\eta_{\mu\nu}+\delta g_{\mu\nu}\,,$$
$$\delta g_{\mu\nu}(x)=-\eta_{\mu\nu}\tfrac{1}{24}R(0)x^2,\quad x^2=(x^l)^2-(x^0)^2\,. \tag{5.10}$$

Varying Eq. (5.7) around a Minkowski background, and dropping terms which are higher than second order in the expansion in powers of $\partial_\lambda g_{\mu\nu}$, we get

$$\delta\left[\frac{1}{8\pi G_{\rm ind}}[R(0)-4\Lambda_{\rm ind}]\right]=\frac{1}{8\pi G_{\rm ind}}R(0)$$
$$=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}\delta\overline{T}[g_{\mu\nu},0]}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}} \quad \text{(i)}$$
$$+\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(0)i\int d^4x\,\delta\overline{\mathscr{L}}}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}} \quad \text{(ii)}$$
$$-\frac{\left[\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(0)\right]\left[\int d\{\phi\}\,e^{iS[\{\phi\},\eta_{\mu\nu}]}i\int d^4x\,\delta\overline{\mathscr{L}}\right]}{\left[\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}\right]^2}\,. \quad \text{(iii)} \tag{5.11}$$

Terms (ii) and (iii) on the right-hand side can be evaluated by using Eqs. (5.6), (5.8), and (5.10), which give

$$i\int d^4x\,\delta\overline{\mathscr{L}}=-\frac{i}{48}R(0)\int d^4x\,x^2T(x)\,. \tag{5.12}$$

To evaluate term (i), we note that since $\delta g_{\mu\nu}$ vanishes as x^2 at $x=0$, the only terms which contribute to $\delta\overline{T}[g_{\mu\nu},0]$ are those in which $\delta g_{\mu\nu}$ is acted on by two derivatives. After a certain amount of algebra, we find

$$\delta\overline{T}[g_{\mu\nu},0]=2R(0)U(0)\,, \tag{5.13}$$

with $U(x)$ the functional defined by

$$U(x)=\tfrac{1}{12}g_{\mu\nu}g_{\alpha\beta}\left[g_{\lambda\theta}\frac{\partial^2\overline{\mathscr{L}}}{\partial(\partial_\lambda g_{\mu\nu})\partial(\partial_\theta g_{\alpha\beta})}-2g_{\theta\phi}\frac{\partial^2\overline{\mathscr{L}}}{\partial g_{\mu\nu}\partial(\partial_\theta\partial_\phi g_{\alpha\beta})}+g_{\theta\phi}\partial_\lambda\frac{\partial^2\overline{\mathscr{L}}}{\partial(\partial_\lambda g_{\mu\nu})\partial(\partial_\theta\partial_\phi g_{\alpha\beta})}\right.$$
$$\left.\left.-2g_{\lambda\theta}\partial_\sigma\frac{\partial^2\overline{\mathscr{L}}}{\partial(\partial_\lambda\partial_\sigma g_{\mu\nu})\partial(\partial_\theta g_{\alpha\beta})}-g_{\theta\phi}\partial_\lambda\partial_\sigma\frac{\partial^2\overline{\mathscr{L}}}{\partial(\partial_\lambda\partial_\sigma g_{\mu\nu})\partial(\partial_\theta\partial_\phi g_{\alpha\beta})}\right]\right|_{g_{\mu\nu}=\eta_{\mu\nu}}. \tag{5.14}$$

Inserting Eqs. (5.12)–(5.14) into Eq. (5.11) and dividing by $2R(0)$ gives the desired formula for $G_{\rm ind}^{-1}$,

$$\frac{1}{16\pi G_{\rm ind}}=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}U(0)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}$$
$$-\frac{i}{96}\int d^4x\,x^2\left[\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(0)T(x)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}-\frac{\left[\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(0)\right]\left[\int d\{\phi\}\,e^{iS[\{\phi\},\eta_{\mu\nu}]}T(x)\right]}{\left[\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}\right]^2}\right]. \tag{5.15}$$

If we define the subtracted functional $\tilde{T}$ by

$$\tilde{T}(x)=T(x)-\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}T(x)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}\,, \tag{5.16}$$

and note that the second term on the right-hand side is a constant, we can rewrite Eq. (5.15) as

$$\frac{1}{16\pi G_{\rm ind}}=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}U(0)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}-\frac{i}{96}\int d^4x\,x^2\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}\tilde{T}(0)\tilde{T}(x)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}\,. \tag{5.17}$$

Finally, recalling the correspondence (Abers and Lee, 1973) between expectations of functionals and vacuum expectations of time-ordered products of the corresponding operators,

$$\langle A(0)\rangle_0=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}A(0)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}\,,$$
$$\langle\mathscr{T}(A(x)B(0))\rangle_0=\frac{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}A(x)B(0)}{\int d\{\phi\}e^{iS[\{\phi\},\eta_{\mu\nu}]}}\,, \tag{5.18}$$

we can rewrite Eqs. (5.9) and (5.15)–(5.17) in the compact form

$$-\frac{1}{2\pi}\frac{\Lambda_{\rm ind}}{G_{\rm ind}}=\langle T(0)\rangle_0\,, \tag{5.19a}$$

$$\frac{1}{16\pi G_{\text{ind}}}=\langle U(0)\rangle_0-\frac{i}{96}\int d^4x\,x^2[\langle\mathscr{T}(T(x)T(0))\rangle_0-\langle T(0)\rangle_0^2]$$

$$=\langle U(0)\rangle_0-\frac{i}{96}\int d^4x\,x^2\langle\mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0\,,$$

$$\tilde{T}(x)=T(x)-\langle T(x)\rangle_0\,. \quad (5.19b)$$

As noted above, we have so far carried along some extra generality, which will be needed to discuss the case when the metric is a quantum variable. When the metric is not quantized, the trace functional $\tilde{T}[g_{\mu\nu},0]$ depends on derivatives of the metric only[20] through terms of order R^2, which come directly from the Lagrangian terms $\mathscr{G},\mathscr{H},\mathscr{K}$ of Eq. (2.35). The variations of these terms vanish in flat space-time, and so when the metric is not quantized, the functional U vanishes. Hence for matter theories on a background manifold, Eq. (5.19b) reduces to the form

$$\frac{1}{16\pi G_{\text{ind}}}=\frac{-i}{96}\int d^4x\,x^2\langle\mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0 \quad (5.20)$$

given by Adler (1980b) and Zee (1981).

B. Convergence and spectral analysis

From the explicit formula of Eq. (5.20), we can again analyze the conditions for G_{ind}^{-1} to be calculable. Since Eq. (5.20) is a flat space-time formula, it will be convenient at this point to switch to the Bjorken-Drell (1965) signature convention, in which Eq. (5.20) becomes

$$\frac{1}{16\pi G_{\text{ind}}}=\frac{i}{96}\int d^4x\,x^2\langle\mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0\,,$$

$$x^2=(x^0)^2-(x^l)^2\,. \quad (5.21)$$

As discussed in Sec. III above, we define the flat space-time matter theory by a renormalization procedure based on dimensional regularization, and so Eq. (5.21) is to be interpreted as a dimensional continuation limit

$$\frac{1}{16\pi G_{\text{ind}}}=\frac{i}{96}\lim_{\omega\to 2}\int d^{2\omega}x\,x^2\langle\mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0^{\omega}\,, \quad (5.22)$$

where $\langle\ \rangle_0^{\omega}$ denotes the vacuum expectation in the 2ω-dimensional theory. Equation (5.22) will give a calculable G_{ind}^{-1} if the integral on the right-hand side is regular at $\omega=2$, and as we have seen, the singularity structure in the ω plane is directly determined by the ultraviolet divergence structure of the dimension-four integral of Eq. (5.21). This can be studied by using the Wilson (1968) operator product expansion of the time-ordered product,[21]

$$\langle\mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0=\frac{\langle\mathscr{O}_0\rangle_0}{(x^2)^4}\times\text{logs}+\frac{\langle\mathscr{O}_2\rangle_0}{(x^2)^3}\times\text{logs}+O\left[\frac{1}{(x^2)^2}\right], \quad (5.23)$$

where "$\times$ logs" indicates the presence of power series in $\log x^2$, and where $\mathscr{O}_{0,2}$ are Lorentz-scalar, internal symmetry-invariant operators of canonical dimension 0 and 2, respectively [corresponding to the fact that $\tilde{T}$ has canonical dimension four, and hence the left-hand side of Eq. (5.23) has canonical dimension eight]. When Eq. (5.23) is inserted in Eq. (5.21) the order $(x^2)^{-4}$ terms give formally quadratically divergent integrals, which vanish by the lemma of Eq. (3.18) above, while the order $(x^2)^{-2}$ and higher terms are ultraviolet convergent. However, the order $(x^2)^{-3}$ terms give logarithmically divergent integrals and thus generate poles at $\omega=2$ in the dimensional continuation, unless no operators $\mathscr{O}_2$ are present in the theory, in which case G_{ind}^{-1} is calculable. We have therefore recovered the same calculability criterion as was obtained from the dimensional algorithm in Sec. II.D above.

Let us next attempt to put Eq. (5.21) into spectral form, which if possible, would yield information about the sign of G_{ind}. From the standard[22] spectral analysis for a scalar operator φ, we have

$$-i\langle\mathscr{T}(\varphi(x)\varphi(0))\rangle_0=\int_0^\infty d\sigma^2\rho(\sigma^2)\Delta_F(x,\sigma)\,, \quad (5.24)$$

with ρ the spectral function defined by[23]

$$\rho(q^2)=(2\pi)^3\sum_n\delta^4(p_n-q)\,|\langle 0|\varphi(0)|n\rangle|^2\geq 0\,, \quad (5.25)$$

and with Δ_F the scalar Feynman propagator,

$$\Delta_F(x,\sigma)=\int\frac{d^4k}{(2\pi)^4}e^{-ik\cdot x}\frac{1}{k^2-\sigma^2+i\epsilon}\,. \quad (5.26)$$

Ignoring for the moment questions of convergence, let us set $\varphi=\tilde{T}$ in the above formulas and substitute into Eq. (5.21), giving

$$\frac{1}{16\pi G_{\text{ind}}}=\frac{1}{96}\int_0^\infty d\sigma^2\rho(\sigma^2)\left[-\int d^4x\,x^2\Delta_F(x,\sigma)\right]. \quad (5.27)$$

A simple calculation then shows that

$$-\int d^4x\,x^2\Delta_F(x,\sigma)=\frac{\partial^2}{\partial k_\mu\partial k^\mu}\frac{1}{(k^2-\sigma^2)}\bigg|_{k=0}=\frac{-8}{\sigma^4}\,, \quad (5.28)$$

and so Eq. (5.27) yields

$$\frac{1}{16\pi G_{\text{ind}}}=-\frac{1}{12}\int_0^\infty d\sigma^2\frac{\rho(\sigma^2)}{\sigma^4}\,, \quad (5.29)$$

which if correct would imply that G_{ind}^{-1} has manifestly the wrong sign to give attractive gravitation. However, Eq. (5.29) is valid only if the integral on the right-hand side converges, which requires the vanishing of $\rho(\sigma^2)/\sigma^2$ as σ becomes infinite. But in gauge theories, we have seen in Sec. III.C above that $\tilde{T}$ contains a trace anomaly

[20]The Lagrangian density $\overline{\mathscr{L}}$ also contains metric derivatives in the spin connections used in constructing the spinor kinetic terms, but these do not appear in the trace functional $\tilde{T}$.

[21]For a proof of the operator product expansion in perturbation theory and a detailed discussion, see Zimmermann (1970).

[22]See Bjorken and Drell (1965), pp. 138—139 and pp. 387—390.

[23]Since ρ is gauge invariant, it can be evaluated in a canonical gauge to establish positivity.

term proportional to the hard operator $[(F^i_{\lambda\sigma})^2]^r$, as a result of which $\rho(\sigma^2)$ behaves asymptotically as $\sigma^4 \times$ logs, invalidating the spectral representation of Eq. (5.29). The failure of the spectral representation, as indicated by the quadratic divergence of Eq. (5.29), is just a reflection of the formal quadratic divergence of Eq. (5.22), arising from the leading $(x^2)^{-4}$ term in the operator product expansion of Eq. (5.23).

The breakdown of Eq. (5.29) can also be rephrased in the language of dispersion relations, by defining

$$\chi(k^2)=\int d^4x\, e^{ik\cdot x}(-i)\langle \mathscr{T}(\widetilde{T}(x)\widetilde{T}(0))\rangle_0 . \qquad (5.30)$$

If $\chi(k^2)-\chi(0)$ obeyed an unsubtracted dispersion relation, then Eq. (5.29) could be derived, but in fact one must make an additional subtraction, as in $\chi(k^2)-\chi(0)-k^2\chi'(0)$, before getting a quantity which obeys an unsubtracted dispersion relation. Substituting this dispersion relation into Eq. (5.21) then yields $G_{\rm ind}^{-1}\propto\chi'(0)$, which furnishes no *a priori* information about the sign of $G_{\rm ind}$. The calculations discussed in the next two sections suggest, in fact, that the sign of $G_{\rm ind}$ is sensitive to details of the infrared behavior of the matter theory.

C. Early model calculations of $G_{\rm ind}^{-1}$

According to Eq. (5.21), the leading perturbative contributions to $G_{\rm ind}^{-1}$ are those in which two insertions of the stress-energy tensor trace T are made in connected matter diagrams of low-loop order, as shown in Fig. 4. In theories with dynamical spontaneous symmetry breaking, such as SU(n) gauge theories, the diagrams of Fig. 4(a) and 4(b) are typically absent and the leading contributions to $G_{\rm ind}^{-1}$ begin at three-loop order. However, one way of simulating the ultraviolet softening produced by dynamical scale-invariance breaking is to consider a massive fermion or scalar meson theory, in which the leading contribution is the one-loop diagram of Fig. 4(a), and to include explicit, finite-mass Pauli-Villars regulators to control the ultraviolet divergences. This calculation has been performed by Sakharov (1975), Akama *et al.*[24] (1978), and Zee (1981), and Zee's results in particular were important in motivating the general derivation leading to Eq. (5.21). Zee considers a fermion loop of mass $m_0=m$, and by including two Pauli-Villars regulators with mass $m_{1,2}$, finds

$$\frac{1}{16\pi G_{\rm ind}}=\frac{2\pi^2}{3(2\pi)^4}I ,$$

$$I=\sum_{i=0}^{2}c_i m_i^2\log m_i^2, \quad \text{with} \sum_{i=0}^{2}c_i=0 , \quad \sum_{i=0}^{2}c_i m_i^2=0 . \quad (5.31)$$

By some simple algebra, Eq. (5.31) can be rewritten as

$$I=m_2^2\left[\frac{m_1^2-m^2}{m_1^2-m_2^2}\right]\log\frac{m_1^2}{m_2^2}-m^2\log\frac{m_1^2}{m_2^2} , \qquad (5.32)$$

an expression which is positive as long as $m^2<m_{1,2}^2$, but which can change sign when the regulator masses are smaller than m, illustrating the sensitivity of the sign of $G_{\rm ind}^{-1}$ to dynamical details. In order to give the observed magnitude of $G_{\rm ind}^{-1}$, Eq. (5.31) requires $m\sim m_{\rm Planck}=1.22\times10^{19}$ GeV, suggesting more generally that to get a realistic theory of Einstein gravitation as an induced quantum effect, the physics of dynamical scale-invariance breaking must take place at energies near the Planck mass.

According to the discussion of Sec. IV.B above, the simplest field theory model which has calculable induced gravitational and cosmological constants is a pure SU(2) gauge theory. A direct evaluation of Eq. (5.7) has been given in this case by Hasslacher and Mottola (1980), using the approximation of saturating the Euclidean continuation of the functional integral by a dilute gas of instantons.[25] Their result can be written as

$$\frac{1}{8\pi G_{\rm ind}}(R-4\Lambda_{\rm ind})+O(R^2)$$
$$=\int_0^{\rho_{\max}(R)}\frac{d\rho}{\rho^5}[C_1+C_2\rho^2R+\cdots]D(\mu\rho) , \qquad (5.33)$$

where the integral is over the instanton size parameter ρ, and where $\rho_{\max}(R)$ symbolically indicates a cutoff on this integration, of unknown form at present, produced by the infrared vacuum structure of the gauge theory. The instanton gas calculation gives a definite expression for the integrand of Eq. (5.33), written as a series expansion in R times the flat space-time instanton density[25] $D(\mu\rho)$,

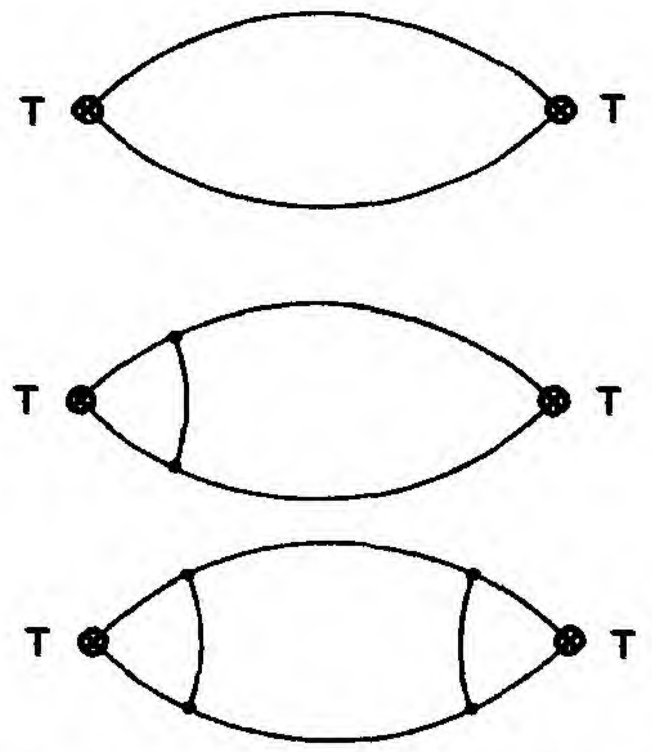


FIG. 4. Typical diagrams contributing to $G_{\rm ind}^{-1}$ in (a) one-, (b) two-, and (c) three-loop order, respectively, with the solid lines indicating matter field propagators. In an SU(n) gauge theory, the contributions of one- and two-loop order vanish, and the perturbation series for $G_{\rm ind}^{-1}$ begins at three-loop order, with a leading term proportional to g^4.

[24]See also Terazawa *et al.* (1977a,b) for related earlier work by this group.

[25]For a pedagogical review of instanton gas methods, see Coleman (1979). A simplified derivation of the instanton density $D(\mu\rho)$ (with μ the subtraction mass discussed in Sec. IV.C) is given by Bernard (1979).

$$\begin{aligned} &C_1=\tfrac{22}{3}\,,\\ &C_2=-\tfrac{5}{3}(\alpha_z+\alpha_\rho+\alpha_g-\tfrac{1}{3}\beta)\,,\\ &\alpha_z=\tfrac{1}{6},\quad \alpha_\rho=\tfrac{1}{3}\log\left[\frac{48}{\rho^2 R}\right]-\frac{7}{24}\,,\\ &\alpha_g=3\alpha_\rho,\quad \beta=\tfrac{1}{6}\,. \end{aligned} \tag{5.34}$$

In Eq. (5.34), C_1 gives the contribution to the cosmological constant arising from the instanton gas expectation of the trace anomaly of Eq. (3.28), while C_2 gives the corresponding contribution to the induced gravitational constant, obtained by summing contributions from the various small fluctuation modes around an instanton. Specifically, α_z, α_ρ, and α_g are, respectively, the contributions from the translational, dilatational and gauge zero modes, while β is the contribution from the nonzero modes. The $\log R$ terms in α_ρ and α_g arise because these zero modes make a contribution to Eq. (5.21) which is infrared divergent. Since an exact evaluation of the Euclidean continuation of the correlation function $\langle \mathscr{T}(\tilde{T}(x)\tilde{T}(0))\rangle_0$ is expected to show an exponential decay law for large separations x (see Sec. V.D below), Eq. (5.21) should in fact be strongly convergent in the infrared. Thus the divergence leading to the presence of $\log R$ in α_ρ and α_g appears to be an artifact of the dilute instanton gas approximation, and one expects the $R\log R$ terms in the integrand of Eq. (5.33) to be cancelled by corresponding terms in the integration cutoff $\rho_{\max}(R)$ and/or in corrections to the instanton picture, leaving a remainder of order R which is determined by the detailed dynamics of the infrared region. This means that the dilute instanton gas calculation, while demonstrating the existence and ultraviolet finiteness of the induced gravitational action in the gauge theory case, does not yield a quantitative calculation of $G_{\rm ind}^{-1}$.

D. A strategy for calculating $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in an SU(n) gauge theory

Because a pure Yang-Mills theory is the simplest field theory model with dynamical scale-invariance breaking, it would clearly be desirable to carry out quantitative calculations of $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in this case. I shall outline below a general strategy for doing this, assuming that one can, in principle, make arbitrarily good Monte Carlo[26] evaluations of the various gluon field vacuum expectations which are needed, together with calculations to any finite order of perturbation theory.

Let us begin with the induced cosmological term $\Lambda_{\rm ind}/G_{\rm ind}$. Substituting Eq. (5.8b) into Eq. (5.19a) and converting to the Bjorken-Drell metric convention (which was used in the derivation of Sec. III.C), we get

$$\frac{1}{2\pi}\frac{\Lambda_{\rm ind}}{G_{\rm ind}}=\langle T_\mu^\mu\rangle_0\,. \tag{5.35}$$

The vacuum expectation on the right can be expressed in terms of the gluon field strength by using the trace anomaly relation of Eq. (3.28), giving

$$\langle T_\mu^\mu\rangle_0=\left\langle\frac{\beta(g)}{2g}(F_{\lambda\sigma}^l F^{l\lambda\sigma})^r\right\rangle_0\,. \tag{5.36}$$

At this point it is convenient to choose a definition of the coupling constant (see Appendix B.1 for details) for which the one-loop renormalization group structure of Eqs. (4.11)–(4.13) is exact.[27] Combining Eq. (5.36) with Eqs. (4.11) and (4.15), we then find

$$\begin{aligned} \frac{1}{2\pi}\frac{\Lambda_{\rm ind}}{G_{\rm ind}}&=\langle T_\mu^\mu\rangle_0\\ &=-\frac{1}{8}\left[\frac{11}{3}n-\frac{2}{3}N_f\right]\left\langle\frac{\alpha_s}{\pi}(F_{\lambda\sigma}^l F^{l\lambda\sigma})^r\right\rangle_0\,,\\ \alpha_s&=\frac{g^2}{4\pi}\,, \end{aligned} \tag{5.37}$$

where for a pure SU(n) Yang-Mills theory one would set $N_f=0$. Equation (5.37) expresses the induced cosmological term as a multiple of the extensively studied[17] gluon pairing amplitude $\langle(\alpha_s/\pi)((F_{\lambda\sigma}^l)^2)^r\rangle_0$. Since the gluon pairing amplitude has canonical dimension four, it is proportional to the fourth power of the renormalization-group-invariant scale mass $\mathscr{M}$ introduced in Sec. IV.B above,

$$\begin{aligned} &\left\langle\frac{\alpha_s}{\pi}(F_{\lambda\sigma}^l F^{l\lambda\sigma})^r\right\rangle_0=c\mathscr{M}^4\,,\\ &\mathscr{M}=\mu e^{-1/(b_0 g^2)}\,, \end{aligned} \tag{5.38}$$

with c a numerical constant of order unity. According to Eq. (5.38), the pairing amplitude has an essential singularity of the form $e^{-4/(b_0 g^2)}$ at $g^2=0$, and vanishes identically in perturbation theory. This agrees with what would be found by making a Feynman diagram expansion of the left-hand side of Eq. (5.38) and evaluating the formally quartically divergent momentum space integrals by using the lemma of Eq. (3.18).

In order to express Eq. (5.38) directly in terms of an observable quantity, it is customary to introduce the string tension σ, defined as the coefficient of the asymptotic linear term in the heavy quark-antiquark static potential,

$$V_{\rm static}(R)\underset{R\to\infty}{=}\sigma R+O(1)\,. \tag{5.39}$$

Since the string tension has canonical dimension two, it is proportional to the square of $\mathscr{M}$,

$$\sigma=c'\mathscr{M}^2\,, \tag{5.40}$$

with c' a second numerical constant of order unity. Eliminating $\mathscr{M}$ from Eqs. (5.38) and (5.40), we get

[26]For a review of statistical physics applications of Monte Carlo methods, see Binder (1976). Lattice gauge theories were introduced by Wilson (1974); see also Kogut and Susskind (1975) and the review by Creutz (1978). The application of Monte Carlo methods to lattice gauge theories was initiated by Creutz, Jacobs, and Rebbi (1979) and Creutz (1980).

[27]If the transformation of Appendix B.1 is not made, the general definition of the gluon pairing amplitude which corresponds to that of Eq. (5.38) is $(-2\beta/b_0 g^3)\langle(\alpha_s/\pi)((F_{\lambda\sigma}^l)^2)^r\rangle_0$.

$$\left\langle \frac{\alpha_s}{\pi}(F^l_{\lambda\sigma}F^{l\lambda\sigma})^r \right\rangle_0 = c''\sigma^2 ,$$
$$c'' = c/(c')^2 , \tag{5.41}$$

which when substituted into Eq. (5.37) gives a relation between the induced cosmological term and the string tension,

$$\frac{\Lambda_{\text{ind}}}{G_{\text{ind}}} = -2\pi\tfrac{1}{8}(\tfrac{11}{3}n - \tfrac{2}{3}N_f)c''\sigma^2 . \tag{5.42}$$

Methods for making a Monte Carlo estimate of c'' in pure SU(2) and SU(3) gauge theories ($n=2,3$; $N_f=0$) have been discussed by Kripfganz (1981), by Banks *et al.* (1981), and by Di Giacomo and Paffuti (1982).

Let us consider next the expression for the induced gravitational constant G_{ind}^{-1} given in Eq. (5.21), which, we have seen, must be interpreted as a dimensional continuation limit. Again substituting the trace anomaly equation, and defining the coupling constant so that the one-loop renormalization group is exact, we get

$$\frac{1}{16\pi G_{\text{ind}}} = \frac{i}{96}\int d^4x\, x^2\Psi(-x^2) ,$$
$$\Psi(-x^2) \equiv \langle \mathscr{T}(T(x)T(0))\rangle_0 - \langle T\rangle_0^2 , \tag{5.43}$$
$$T = -\tfrac{1}{4}b_0 g^2 (F^l_{\lambda\sigma}F^{l\lambda\sigma})^r .$$

To evaluate Eq. (5.43) it is convenient to make a Wick rotation to the Euclidean section, which is formally accomplished by making the substitutions $d^4x \to -i\,d^4x$, $x^2 \to -x^2$, giving

$$\frac{1}{16\pi G_{\text{ind}}} = -\frac{1}{96}\int_E d^4x\, x^2\Psi(x^2) . \tag{5.44}$$

In order to devise a practical method for implementing the dimensional continuation limit implicit in Eq. (5.44),[28] we shall split the integration over the variable $x^2=t$ into an ultraviolet (UV) part $0\le t\le t_0$, and an infrared (IR) part $t_0 \le t < \infty$,

$$\frac{1}{16\pi G_{\text{ind}}} = -\frac{\pi^2}{96}(I_{\text{UV}} + I_{\text{IR}}) ,$$
$$I_{\text{UV}} = \int_0^{t_0} dt\, t^2\Psi(t) , \tag{5.45}$$
$$I_{\text{IR}} = \int_{t_0}^{\infty} dt\, t^2\Psi(t) .$$

Let us suppose that the correlation function $\Psi(t)$ has been determined to high accuracy by Monte Carlo studies. In the infrared region, Ψ behaves for large t as

$$\Psi(t) \underset{t\to\infty}{\sim} e^{-m_g t^{1/2}} , \tag{5.46}$$

with m_g a parameter, called the glueball mass, which is related to the string tension by

$$m_g = c_g\sigma^{1/2} = c_g(c')^{1/2}\mathscr{M} , \tag{5.47}$$

with c_g a numerical constant. [Numerical Monte Carlo estimates of c_g for an SU(2) gauge theory, obtained by studying the plaquette-plaquette correlation function, have been given recently by Berg (1981) and by Bhanot and Rebbi (1981).] As a result of the good asymptotic behavior of Eq. (5.46), the infrared integral I_{IR} of Eq. (5.45) is convergent at $t=\infty$, and can be evaluated by numerical integration. Turning next to the ultraviolet integral I_{UV}, let us write it in the form

$$I_{\text{UV}} = I^c_{\text{UV}} + \Delta I_{\text{UV}} ,$$
$$I^c_{\text{UV}} = \int_0^{t_0} dt\, t^2\Psi_c(t) ,$$
$$\Delta I_{\text{UV}} = \int_0^{t_0} dt\, t^2[\Psi(t) - \Psi_c(t)] , \tag{5.48}$$

with $\Psi_c(t)$ a comparison function chosen so that: (i) the integral ΔI_{UV} converges at $t=0$, and hence can be evaluated by numerical integration; and (ii) the dimensional continuation needed to evaluate I^c_{UV} can be carried out explicitly, leaving a convergent integral which can again be done numerically. The motivation behind the introduction of Ψ_c is the evident fact that, while discrete methods can be used to evaluate convergent integrals, they cannot be used to make analytic continuations.

The general form required for the comparison function $\Psi_c(t)$ can be inferred from the operator product expansion of Eq. (5.23). This expansion can be "improved" by using the renormalization group and asymptotic freedom, which permit a partial resummation of the power series of logarithms in the leading term of Eq. (5.23) into a joint power series in the running coupling constant $g^2(t)$ and (since we have made the transformation of Appendix B.1) in its logarithm $\log g^2(t)$. Defining the coordinate space running coupling by[29]

$$g^2(t) \equiv \frac{1}{-\frac{1}{2}b_0\log(\mathscr{M}^2 t)} = \frac{g^2}{1 - \frac{1}{2}b_0 g^2\log(\mu^2 t)} , \tag{5.49}$$

we have[30]

[28]Use of a coordinate space formalism is not necessary in order to implement the dimensional continuation limit. For example, one could equally well rewrite the spectral representation of Eq. (5.29) as

$$\frac{1}{16\pi G_{\text{ind}}} = -\frac{1}{12}(J_{\text{UV}} + J_{\text{IR}}) ,$$
$$J_{\text{UV}} = \int_{\sigma_0^2}^{\infty} d\sigma^2 \frac{\rho(\sigma^2)}{\sigma^4},\quad J_{\text{IR}} = \int_0^{\sigma_0^2} d\sigma^2\frac{\rho(\sigma^2)}{\sigma^4} ,$$

and evaluate J_{UV} by dimensional continuation. However, it is likely to be easier to extract the coordinate space function $\Psi(x^2)$ than the spectral function $\rho(\sigma^2)$ from Monte Carlo studies of the infrared region.

[29]The use of the same scale mass in Eq. (5.49) as in the one-loop version of Eq. (4.10) is a matter of convenience; redefining $\mathscr{M}$ by a constant factor simply redefines the expansion coefficients appearing in Eq. (5.50).

[30]In general, such renormalization-group-improved operator product expansions contain an additional fractional power $[\log(\mathscr{M}^2 t)]^\delta$, with the exponent δ proportional to the difference in anomalous dimensions of the operators on the left- and right-hand sides. Since T^μ_μ and $\mathscr{O}_0 \propto 1$ both have zero anomalous dimension, this fractional power is absent from the leading term in the expansion. See Gross and Wilczek (1974), p. 982, for a detailed discussion of this point.

$$\Psi(t)=C_\Psi \frac{1}{t^4(-\log \mathscr{M}^2 t)^2}\left[1+\frac{1}{(-\log \mathscr{M}^2 t)}[a_{10}+a_{11}\log\log(\mathscr{M}^2 t)^{-1}]+\cdots\right]+O(t^{-2}) . \tag{5.50}$$

The leading term in Eq. (5.50) is proportional to

$$\frac{1}{(-\log \mathscr{M}^2 t)^2}\propto g^4(t) , \tag{5.51}$$

because, as seen from Eq. (5.43), the perturbation expansion for Ψ begins in order g^4; the constant C_Ψ is computed from lowest-order perturbation theory in Appendix B.2, with the result

$$C_\Psi=\frac{3\times 2^6}{(2\pi)^4}(n^2-1) . \tag{5.52}$$

[The two-loop contribution to the glueball propagator, which gives the coefficients a_{10},a_{11} in the series of Eq. (5.50), has recently been calculated by Kataev *et al.* (1982).] No order t^{-3} term is present in the expansion of Eq. (5.50) because of the absence of dimension-two operators $\mathscr{O}_2$, while the order t^{-2} and higher terms make contributions to $I_{\rm UV}$ which are convergent at $t=0$. Hence it suffices to take as the comparison function Ψ_c the leading t^{-4} part of $\Psi(t)$,

$$\Psi_c(t)= C_\Psi \frac{1}{t^4(-\log \mathscr{M}^2 t)^2} \times\left[1+\sum_{n=1}^{\infty}\sum_{m=0}^{n} a_{nm}\frac{[\log\log(\mathscr{M}^2 t)^{-1}]^m}{(-\log \mathscr{M}^2 t)^n}\right], \tag{5.53}$$

and to restrict t_0 by the condition

$$\mathscr{M}^2 t_0<1 , \tag{5.54}$$

so that the logarithm $\log(\mathscr{M}^2 t)$ does not vanish in the integration range $0\le t\le t_0$ of $I^c_{\rm UV}$. Substituting Eq. (5.53) into $I^c_{\rm UV}$ and making the change of variable $u=\mathscr{M}^2 t$ gives

$$I^c_{\rm UV}=C_\Psi \mathscr{M}^2\int_0^{u_0}\frac{du}{u^2}\frac{\Theta(u)}{(\log u)^2},\quad u_0=\mathscr{M}^2 t_0 ,$$

$$\Theta(u)=1+\sum_{n=1}^{\infty}\sum_{m=0}^{n} a_{nm}\frac{(\log\log u^{-1})^m}{(\log u^{-1})^n} . \tag{5.55}$$

The evaluation of this integral by dimensional continuation is carried out in Appendix B.3, with the result

$$I^c_{\rm UV}=C_\Psi \mathscr{M}^2 {\rm Re}\left[\int_{\log(\mathscr{M}^2 t_0)^{-1}}^{i\infty} dv\frac{e^v}{v^2}\Theta(e^{-v})\right], \tag{5.56}$$

where Re indicates the real part, and where the integration contour is shown in Fig. 5. [As discussed in Appendix B.3, the need to take a real part in Eq. (5.56), reflecting the existence of a cut in the ω plane, arises from the fact that the running coupling constant variable $g^2(t)$ used in the "improved" expansion sums an infinite number of Feynman diagrams. The dimensional continuation of individual Feynman diagrams remains meromorphic in ω.] The integral of Eq. (5.56) can be done by numerical integration, and so the problem of evaluating Eq. (5.43) has been reduced to a sequence of steps which can each be implemented by discrete methods.

Up to this point in the discussion I have used the one-loop exact running coupling constant defined in Eq. (5.49), which transforms the renormalization group to its minimal, exponential form. However, in doing an actual calculation it is not advantageous to make the nonanalytic transformation of Appendix B.1; instead, it is better to work with a two-loop exact or more general definition of the running coupling constant $g^2(t)$, in terms of which $\Psi_c(t)$ takes the form of a simple power-series expansion

$$\Psi_c(t)=\tfrac{1}{4}b_0^2 C_\Psi \frac{[g^2(t)]^2}{t^4}\left[1+\sum_{n=1}^{\infty} c_n[g^2(t)]^n\right] . \tag{5.57}$$

Corresponding to this, Eqs. (5.55) and (5.56) take the form

$$\Theta(u)=1+\sum_{n=1}^{\infty} c_n[g^2(u/\mathscr{M}^2)]^n ,$$

$$I^c_{\rm UV}=\tfrac{1}{4}b_0^2 C_\Psi \mathscr{M}^2{\rm Re}\left\{\int_{\log(\mathscr{M}^2 t_0)^{-1}}^{i\infty} dv\, e^v \times[g^2(e^{-v}/\mathscr{M}^2)]^2\Theta(e^{-v})\right\}, \tag{5.58}$$

with the coefficient c_1 known from the above-cited work of Kataev *et al.*, and with the higher coefficients yet to be computed. In doing a calculation it is of course necessary to make an explicit choice both for the dividing point t_0, and for the accuracy to which the perturbation expansion Ψ_c is to be computed. A reasonable strategy for doing this, I believe, is as follows:

(i) Choose t_0 far enough into the ultraviolet so that perturbation theory is valid at t_0, and so that $|\Delta I_{\rm UV}/I_{\rm IR}|$ is small. Such a choice is always possible, since the fact that $\Delta I_{\rm UV}$ is a convergent integral implies that

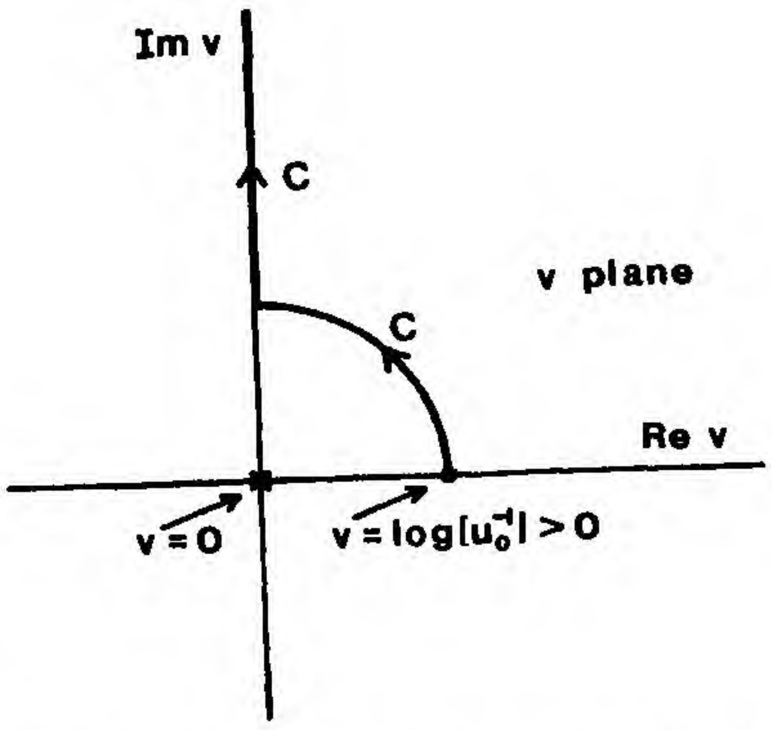


FIG. 5. Contour of integration C to be used in evaluating Eq. (5.56). The contour begins at $v=\log u_0^{-1}=\log(\mathscr{M}^2 t_0)^{-1}$ and must avoid the singularity at $v=0$.

$$\lim_{t_0\to 0} \Delta I_{UV}=0 . \tag{5.59}$$

(ii) Then, keeping t_0 fixed, compute a large enough number N of perturbation-theory coefficients c_n so that I^c_{UV} is well approximated by

$$I^{cN}_{UV}=\left[\int_0^{t_0} dt\, t^2\Psi^N_c(t)\right]_{\text{dimensionally regularized}}$$
$$=\tfrac{1}{4}b_0^2C_\Psi\mathcal{M}^2\text{Re}\left\{\int_{\log(\mathcal{M}^2t_0)^{-1}}^{i\infty} dv\, e^v \times[g^2(e^{-v}/\mathcal{M}^2)]^2\Theta^N(e^{-v})\right\}, \tag{5.60}$$

with $\Psi^N_c(t)$ and $\Theta^N(u)$, respectively, the truncations of the series of Eq. (5.57) and Eq. (5.58) to the first N terms. Such an approximation is possible because[31]

$$\lim_{N\to\infty} \Psi^N_c(t)=\Psi_c(t) \tag{5.61}$$

implies that

$$\lim_{N\to\infty} I^{cN}_{UV}=I^c_{UV} . \tag{5.62}$$

(iii) According to Eqs. (5.59) and (5.62), the total integral which we are calculating is given by the double limit

$$I\equiv I_{IR}+I^c_{UV}+\Delta I_{UV}$$
$$=\lim_{t_0\to 0}\lim_{N\to\infty}(I_{IR}+I^{cN}_{UV}) , \tag{5.63}$$

which with t_0 and N chosen according to (i) and (ii), is well approximated by

$$I\approx I_{IR}+I^{cN}_{UV} . \tag{5.64}$$

However, for fixed N we must be careful not to let t_0 become arbitrarily small in the approximated expression of Eq. (5.64), because as a result of the mismatch between I_{IR} and I^{cN}_{UV} and the quadratic divergence of the unregularized integral, we find

$$\lim_{t_0\to 0}(I_{IR}+I^{cN}_{UV})=\infty . \tag{5.65}$$

In other words, the order of the limiting operations in Eq. (5.63) is significant, and is reflected in the procedure for choosing t_0 and N given in (i) and (ii) above.

In a recent paper, Zee (1982a) has given a model in which the infrared region is explicitly known, permitting the complete integral $I_{UV}+I_{IR}$ to be evaluated explicitly by dimensional regularization, and thus giving a simple illustration of the methods outlined above. Zee's model is a gauge theory in which the one-loop β-function coefficient b_0 is positive and small, while the two-loop β-function coefficient b_1 is negative [cf. Appendix B, Eq. (B1)], as happens, for instance, in QCD with 16 quark flavors. Such a theory is still asymptotically free, but has a nontrivial infrared stable fixed point at a small coupling constant $g^2_*=-b_0/(2b_1)$. In the approximation of retaining only the leading term in an expansion in powers of g^2_*, one finds

$$\Psi(t)=\Psi_c(t)=\tfrac{1}{4}b_0^2C_\Psi\frac{1}{t^4}\left[g^2(t)\left[1-\frac{g^2(t)}{g^2_*}\right]\right]^2, \tag{5.66}$$

with the two factors $g^2(t)[1-g^2(t)/g^2_*]$ arising directly from the two factors $\beta(g)/g$, which appear in Ψ when the trace anomaly formula of Eq. (3.28) is used. Hence, in this model the entire answer is given by the power-series expansion of Eq. (5.57), and the series terminates after only a finite number of terms. The explicit calculation shows that the sign of G_{ind} in this model depends strongly on the values of the β-function coefficients b_0 and b_1, and thus again is sensitive to infrared details. [For a further discussion, in the context of a survey of induced gravitation generally, see Zee (1982c).]

VI. EXTENSION TO A QUANTIZED METRIC

A. The general-coordinate invariant effective action, and derivation of the background metric Einstein equations

Up to this point the metric $g_{\mu\nu}$ has been treated as a purely classical variable, which determines the background geometry and thereby influences the dynamics of the quantized matter fields, but which is not itself quantized. While this classical metric formulation is useful as a model, there are a number of arguments indicating that it is not a satisfactory starting point for a fundamental theory. For example, Duff (1981) has pointed out that if the metric is not quantized, then the system of equations comprising the quantized matter fields and the classical Einstein equations for the metric is not invariant under metric-dependent redefinitions of the matter fields. Such redefinitions should be allowed in a completely consistent formulation, and Duff shows that they are in fact permitted if the metric is quantized. A second argument is simply that if the metric is treated as a classical variable, then the Einstein equations or the equivalent Einstein-Hilbert action principle must be postulated on an *ad hoc* basis. As we will see below, when the metric is quantized, the Einstein equations for the background metric emerge automatically as the leading long-wavelength approximation to the effective action formalism.

In discussing the dynamics of a quantized matter-metric system, it is necessary to give a procedure for identifying that part $\bar g_{\mu\nu}$ of the metric which we observe as the "classical" metric and a method for computing its effective action functional. I do this by using the background field method of DeWitt (1965), in which the total quantum metric $g_{\mu\nu}$ is split, in a self-consistent fashion, into the sum of a background metric $\bar g_{\mu\nu}$ and a quantum fluctuation $h_{\mu\nu}$,

[31]If the series for $\Psi_c(t)$ is only an asymptotic series, a summation procedure [such as Padé approximants or Borel summation; see Simon (1981)] is needed to extract, from the perturbation coefficients c_n, a sequence of approximants Ψ^N_c which satisfy Eq. (5.61).

$$g_{\mu\nu}=\bar{g}_{\mu\nu}+h_{\mu\nu} . \tag{6.1}$$

Elaborating on earlier work by 't Hooft (1975), recently Boulware (1981) and DeWitt (1981) have given an extension[32] of the background field method which preserves manifest general-coordinate covariance with respect to the background metric, and hence is an ideal vehicle for the discussion which follows.

To introduce the general-coordinate invariant effective action formalism, let us consider first the case in which no matter fields are present, so that the total microscopic action density consists solely of the term $\mathscr{L}_{\rm grav}[g_{\mu\nu}]$ introduced in Eq. (2.38). The partition function is then given formally by

$$Z=\int d[g_{\mu\nu}]e^{iS_{\rm grav}[g_{\mu\nu}]} ,$$

$$S_{\rm grav}[g_{\mu\nu}]=\int d^4x \sqrt{-g}\,\mathscr{L}_{\rm grav}[g_{\mu\nu}] , \tag{6.2}$$

but this expression is divergent because of the general-coordinate invariance of the action. To get a useful expression for Z, a gauge-fixing term and a compensating Fadde'ev-Popov (1967) determinant must be introduced into Eq. (6.2). Let us choose the gauge-fixing term in the action to have the form

$$S_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]=\int d^4x \sqrt{-g^R}\,\mathscr{L}_{gf}[g^R_{\alpha\beta},g_{\mu\nu}] , \tag{6.3}$$

with $g^R_{\alpha\beta}$ an arbitrary fixed reference metric (which for the time being is distinct from $\bar{g}_{\mu\nu}$), and with $\mathscr{L}_{gf}$ constructed so as to transform formally as a general-coordinate scalar with respect to $g^R_{\alpha\beta}$, when the total quantum metric $g_{\mu\nu}$ is treated as a tensor with respect to $g^R_{\alpha\beta}$. A suitable gauge fixing for quantizing the curvature-squared action of Eq. (2.38) would be[33]

$$\mathscr{L}_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]=\tfrac{1}{2}g^{R\lambda\sigma}g^{R\mu\nu}\nabla_{R\lambda}G_\mu\nabla_{R\sigma}G_\nu , \tag{6.4}$$

with ∇_R the covariant derivative with respect to $g^R_{\mu\nu}$, and with G_ν formally a covariant vector with respect to $g^R_{\mu\nu}$, given by

$$G_\nu=\nabla^\mu_R g_{\mu\nu}-\tfrac{1}{2}g^{R\mu\lambda}\nabla_{R\nu}g_{\mu\lambda} . \tag{6.5}$$

Equations (6.4) and (6.5) are a natural generalization of the usual harmonic coordinate condition; however, the precise form of $\mathscr{L}_{gf}$ (beyond the fact that it depends explicitly on the auxiliary metric $g^R_{\alpha\beta}$) will not play a role in the following discussion. The gauge fixing of Eqs. (6.4) and (6.5) completely breaks the invariance of the gravitational action under the group of general-coordinate transformations $g_{\mu\nu}\to g^\theta_{\mu\nu}$, which has the infinitesimal form[34]

$$\delta_\theta g_{\mu\nu}\equiv g_{\mu\lambda}\partial_\nu(\delta\theta^\lambda)+g_{\lambda\nu}\partial_\mu(\delta\theta^\lambda)+(\partial_\lambda g_{\mu\nu})\delta\theta^\lambda , \tag{6.6}$$

with $\delta\theta^\lambda$ an arbitrary infinitesimal contravariant vector. The Fadde'ev-Popov compensating determinant for the gauge-fixing action of Eq. (6.3) is defined by[35]

$$1=\int d[\theta]e^{iS_{gf}[g^R_{\alpha\beta},g^\theta_{\mu\nu}]}\Delta[g^R_{\alpha\beta},g_{\mu\nu}] , \tag{6.7}$$

with $d[\theta]$ the invariant measure on the manifold of the general-coordinate transformation group. Since the invariant measure satisfies

$$d[\theta\theta']=d[\theta'\theta]=d[\theta] \tag{6.8}$$

for any fixed general-coordinate transformation $g_{\mu\nu}\to g^{\theta'}_{\mu\nu}$, we learn from Eqs. (6.7) and (6.8) that the compensating determinant is invariant under general-coordinate transformations on $g_{\mu\nu}$,

$$\Delta[g^R_{\alpha\beta},g_{\mu\nu}]=\Delta[g^R_{\alpha\beta},g^{\theta'}_{\mu\nu}] . \tag{6.9}$$

According to the Fadde'ev-Popov ansatz,[35] a convergent path-integral representation for the partition function is then given by

$$Z=\int d[g_{\mu\nu}]\Delta[g^R_{\alpha\beta},g_{\mu\nu}]e^{iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]} . \tag{6.10}$$

---

[32]See also Fradkin and Vilkovisky (1976), who use the gauge fixing

$$\mathscr{L}_{gf}=\tfrac{1}{2}g^{R\mu\nu}G_\mu G_\nu \,|_{g^R_{\mu\nu}=\bar{g}_{\mu\nu}}$$

to quantize the Einstein theory formally, and who suggest that it gives a generally covariant effective action for $\bar{g}_{\mu\nu}$. For the use of the gauge-invariant background field method to compute two-loop counter terms, see Abbott (1981) and Ichinose and Omote (1982).

[33]For a discussion of the complexities involved in representing higher-derivative gauge fixings in terms of a local "ghost" action density, see Kallosh (1978) and Nielsen (1978).

[34]As pointed out by Fradkin and Vilkovisky (1975) and reviewed by Batalin and Fradkin (1979), the presence of a term $\partial_\lambda g_{\mu\nu}$ in $\delta_\theta g_{\mu\nu}$ leads to a nonvanishing variation of the integration measure under general coordinate transformations,

$$\delta_\theta d[g_{\mu\nu}]\propto {\rm Tr}[\delta(\delta_\theta g_{\mu\nu}(x))/\delta g_{\lambda\sigma}(y)]$$
$$\propto \int d^4x\,\partial_\lambda\delta^4(0)\delta\theta^\lambda(x) .$$

The $\partial_\lambda\delta^4(0)$ term vanishes in covariant calculations using dimensional regularization, and is ignored in the discussion of the text, where $d[g_{\mu\nu}]$ is treated as being general-coordinate invariant. The variation of the integration measure cannot be ignored in setting up a canonical, Hamiltonian formalism using a massive regulator scheme; in this case it leads to an extra Jacobian factor in the path-integral formulas, which can be represented by a quartic local "ghost" action density. For a related analysis of the connection between Jacobian factors in the path-integral measure and chiral and conformal anomalies, see Fujikawa (1981).

[35]The discussion of Eqs. (6.7)—(6.14) is based on Sec. 3.3 of Fadde'ev and Slavnov (1980). Strictly speaking, the Lagrangian form of the path-integral formula given in Eq. (6.10) must be derived from the more fundamental Hamiltonian form, and the standard textbook discussions describe this step only for second-order actions. The derivation of Eq. (6.10) from the Hamiltonian formalism in the case of fourth-order, curvature-squared gravitational actions has been carried out by Boulware (1982).

To verify that Z is independent of the choice of the reference metric $g^R_{\alpha\beta}$, let us multiply the integrand of Eq. (6.10) by unity in the form

$$1=\int d[\theta]e^{iS_{gf}[g'^R_{\alpha\beta},g^\theta_{\mu\nu}]}\Delta[g'^R_{\alpha\beta},g^\theta_{\mu\nu}]\ , \tag{6.11}$$

giving

$$Z=\int d[\theta]d[g_{\mu\nu}]\Delta[g^R_{\alpha\beta},g_{\mu\nu}]\Delta[g'^R_{\alpha\beta},g^\theta_{\mu\nu}] \times e^{iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]+iS_{gf}[g'^R_{\alpha\beta},g^\theta_{\mu\nu}]}\ . \tag{6.12}$$

Making the substitution $g_{\mu\nu}\to g^{\theta^{-1}}_{\mu\nu}$, and using the fact that the action $S_{\rm grav}[g_{\mu\nu}]$, the compensating determinants Δ and the integration measure[34] $d[g_{\mu\nu}]$ are all general-coordinate invariant, and also using the invariance property $d[\theta]=d[\theta^{-1}]$, Eq. (6.12) becomes

$$Z=\int d[\theta^{-1}]d[g_{\mu\nu}]\Delta[g^R_{\alpha\beta},g^{\theta^{-1}}_{\mu\nu}]\Delta[g'^R_{\alpha\beta},g_{\mu\nu}] \times e^{iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g^{\theta^{-1}}_{\mu\nu}]+iS_{gf}[g'^R_{\alpha\beta},g_{\mu\nu}]}\ . \tag{6.13}$$

But now applying Eq. (6.7) once more (with θ replaced by θ^{-1}) we get

$$Z=\int d[g_{\mu\nu}]\Delta[g'^R_{\alpha\beta},g_{\mu\nu}]e^{iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g'^R_{\alpha\beta},g_{\mu\nu}]}\ , \tag{6.14}$$

which differs from the original form in Eq. (6.10) by the replacement of $g^R_{\alpha\beta}$ by $g'^R_{\alpha\beta}$.

Let us now introduce an external source $J^{\lambda\sigma}$ coupled to the metric $g_{\lambda\sigma}$, so that the path-integral formula of Eq. (6.10) is modified to read

$$e^{iW[J^{\lambda\sigma},g^R_{\alpha\beta}]}\equiv Z[J^{\lambda\sigma},g^R_{\alpha\beta}] =\int d[g_{\mu\nu}]\Delta[g^R_{\alpha\beta},g_{\mu\nu}] \times e^{iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]-i\int d^4x\, g_{\lambda\sigma}J^{\lambda\sigma}}\ . \tag{6.15}$$

Both $J^{\lambda\sigma}$ and $g^R_{\alpha\beta}$ are indicated as arguments of Z in Eq. (6.15) because the source term breaks the general-coordinate invariance of the action. As a result, when $J^{\lambda\sigma}\neq 0$ the argument of Eqs. (6.11)–(6.14) cannot be applied, and hence the previously derived zero-source invariance,

$$0=\frac{\delta}{\delta g^R_{\alpha\beta}}Z[0,g^R_{\alpha\beta}]=\frac{\delta}{\delta g^R_{\alpha\beta}}W[0,g^R_{\alpha\beta}]\ , \tag{6.16}$$

cannot be extended to the case when a source is present. From the functional W, we can calculate the expectation value $\bar g_{\lambda\sigma}$ of the metric in the presence of the source $J^{\lambda\sigma}$ by using the formula

$$\bar g_{\lambda\sigma}[J^{\gamma\delta},g^R_{\alpha\beta}]=\langle g_{\lambda\sigma}\rangle_J=-\frac{\delta W}{\delta J^{\lambda\sigma}}\ , \tag{6.17}$$

which can be inverted to determine $J^{\lambda\sigma}$ implicitly as a functional of $\bar g_{\lambda\sigma}$ (and of $g^R_{\alpha\beta}$),

$$J^{\lambda\sigma}=J^{\lambda\sigma}[\bar g^{\gamma\delta},g^R_{\alpha\beta}]\ . \tag{6.18}$$

Let us now introduce the Legendre-transformed effective action functional Γ defined by

$$\Gamma=W+\int d^4x\, \bar g_{\lambda\sigma}J^{\lambda\sigma}\ . \tag{6.19}$$

Varying Eq. (6.19) (for fixed $g^R_{\alpha\beta}$) and using Eq. (6.17), we get

$$\delta\Gamma=\delta W+\int d^4x(\bar g_{\lambda\sigma}\delta J^{\lambda\sigma}+\delta\bar g_{\lambda\sigma}J^{\lambda\sigma}) =\int d^4x\, \delta\bar g_{\lambda\sigma}J^{\lambda\sigma}\ , \tag{6.20}$$

which shows that Γ is a functional only of $\bar g^{\lambda\sigma}$ and $g^R_{\alpha\beta}$,

$$\Gamma=\Gamma[\bar g^{\lambda\sigma},g^R_{\alpha\beta}]\ , \tag{6.21a}$$

and satisfies

$$\frac{\delta\Gamma}{\delta\bar g_{\lambda\sigma}}=J^{\lambda\sigma}\ . \tag{6.21b}$$

The partition function $Z[J^{\lambda\sigma},g^R_{\alpha\beta}]$ can be reexpressed in terms of the effective action Γ through the formula

$$e^{iW[J^{\lambda\sigma},g^R_{\alpha\beta}]}={\rm ext}_{g^{\lambda\sigma}}\left[e^{i\Gamma[g^{\lambda\sigma},g^R_{\alpha\beta}]-i\int d^4x\, g_{\lambda\sigma}J^{\lambda\sigma}}\right]\ , \tag{6.22}$$

where ${\rm ext}_{g^{\lambda\sigma}}(\ )$ indicates that one is to take the extremum of the parenthesis over all values of $g^{\lambda\sigma}$. Equation (6.22) is verified by noting that the exponent on the right-hand side is extremized[36] at the metric $g^{\lambda\sigma}=\bar g^{\lambda\sigma}$ for which Eq. (6.21b) is satisfied, and that at the extremum it can be rewritten, by using Eq. (6.19), to give $W[J^{\lambda\sigma},g^R_{\alpha\beta}]$.

With these preliminaries completed, we are ready to introduce the general-coordinate invariant effective action functional $\Gamma_{\rm inv}[\bar g^{\lambda\sigma}]$, defined by identifying the reference metric $g^R_{\alpha\beta}$ with the expectation value $\bar g_{\alpha\beta}$ in the formulas given above,

$$\Gamma_{\rm inv}[\bar g^{\lambda\sigma}]\equiv\Gamma[\bar g^{\lambda\sigma},\bar g_{\alpha\beta}]\ . \tag{6.23}$$

To get an explicit formula for $\Gamma_{\rm inv}$, let us multiply Eq. (6.15) by $\exp(i\int d^4x\, \bar g_{\lambda\sigma}J^{\lambda\sigma})$ and change to $h_{\mu\nu}$, defined in Eq. (6.1), as the new functional integration variable. Making use of the identity

$$S_{gf}[\bar g_{\alpha\beta},\bar g_{\mu\nu}+h_{\mu\nu}]=S_{gf}[\bar g_{\alpha\beta},h_{\mu\nu}] =\tfrac{1}{2}\bar g^{\lambda\sigma}\bar g^{\mu\nu}\bar\nabla_\lambda G_\mu\bar\nabla_\sigma G_\nu\ ,$$

$$G_\nu=\bar\nabla^\mu h_{\mu\nu}-\tfrac{1}{2}\bar g^{\mu\lambda}\bar\nabla_\nu h_{\mu\lambda} \tag{6.24}$$

(which follows from the fact that $\bar\nabla_\lambda\bar g_{\mu\nu}=0$), we get the following functional integral representation for $\Gamma_{\rm inv}$,

$$e^{i\Gamma_{\rm inv}[\bar g^{\lambda\sigma}]} =\int d[h_{\mu\nu}]\Delta[\bar g_{\alpha\beta},\bar g_{\mu\nu}+h_{\mu\nu}] \times e^{iS_{\rm grav}[\bar g_{\mu\nu}+h_{\mu\nu}]+iS_{gf}[\bar g_{\alpha\beta},h_{\mu\nu}]-i\int d^4x\, h_{\lambda\sigma}J^{\lambda\sigma}[\bar g_{\alpha\beta}]}\ . \tag{6.25}$$

---

[36] I will assume here, and later on, that the extremum problems which are encountered always have a unique solution.

The source current $J^{\lambda\sigma}$ in Eq. (6.25) is implicitly determined as a functional of $\bar{g}^{\lambda\sigma}$ by the requirement

$$0=\langle h_{\lambda\sigma}\rangle_J=-\frac{\delta\Gamma}{\delta J^{\lambda\sigma}}\,, \tag{6.26}$$

which is equivalent to

$$0=\int d[h_{\mu\nu}]\Delta[\bar{g}_{\alpha\beta},\bar{g}_{\mu\nu}+h_{\mu\nu}]h_{\lambda\sigma} \times e^{iS_{\rm grav}[\bar{g}_{\mu\nu}+h_{\mu\nu}]+iS_{gf}[\bar{g}_{\alpha\beta},h_{\mu\nu}]-i\int d^4x\, h_{\xi\eta}J^{\xi\eta}[\bar{g}_{\alpha\beta}]}\,. \tag{6.27}$$

To see that $\Gamma_{\rm inv}$ is a general-coordinate invariant functional of its argument, we note that we are free to take $h_{\mu\nu}$, which is a dummy integration variable, to transform as a tensor with respect to general-coordinate transformations of $\bar{g}_{\mu\nu}$. By construction, S_{gf} is then a scalar with respect to such transformations, and therefore from Eq. (6.7), the compensating determinant $\Delta[\bar{g}_{\alpha\beta},\bar{g}_{\mu\nu}+h_{\mu\nu}]$ is also a scalar. Equation (6.27) then determines $J^{\lambda\sigma}$ to transform as a tensor, and so the right-hand side of Eq. (6.25) is manifestly invariant under general-coordinate transformations of $\bar{g}_{\mu\nu}$.

Let us next show that the source-free partition function Z can be obtained by extremizing the gauge-invariant effective action functional. According to Eqs. (6.22) and (6.16), in the absence of an external source we have

$$Z=\operatorname{ext}_{g^{\lambda\sigma}}(e^{i\Gamma[g^{\lambda\sigma},g^R_{\alpha\beta}]})\,, \tag{6.28a}$$

$$\frac{\delta}{\delta g^R_{\alpha\beta}}Z=0\,. \tag{6.28b}$$

The extremum in Eq. (6.28a) determines $g^{\lambda\sigma}$ to take a value $\bar{g}^{\lambda\sigma}\,[g^R_{\alpha\beta}]$ at which

$$\frac{\delta\Gamma}{\delta g^{\lambda\sigma}}[\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]=0\,, \tag{6.29}$$

and when expressed in terms of $\bar{g}^{\lambda\sigma}$, the reference-metric invariance of Eq. (6.28b) takes the form

$$\frac{\delta\Gamma}{\delta g^R_{\alpha\beta}}[\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]=0\,. \tag{6.30}$$

Since $\bar{g}^{\lambda\sigma}[g^R_{\alpha\beta}]$ is a continuous map from the manifold of reference metrics into itself, there is[37] a fixed point $g^R_{\alpha\beta}=g^*_{\alpha\beta}$ for which $\bar{g}^{\lambda\sigma}[g^*_{\alpha\beta}]=g^{*\lambda\sigma}$. At the fixed point, Eqs. (6.29) and (6.30) become

$$\frac{\delta}{\delta g^{\lambda\sigma}}\Gamma[g^{*\lambda\sigma},g^*_{\alpha\beta}]=\frac{\delta}{\delta g^R_{\alpha\beta}}\Gamma[g^{*\lambda\sigma},g^*_{\alpha\beta}]=0\,, \tag{6.31}$$

which together imply that

$$\frac{\delta}{\delta g^{\lambda\sigma}}\Gamma_{\rm inv}[g^{*\lambda\sigma}]=0\,, \tag{6.32}$$

and so[36] we have

$$Z=\operatorname{ext}_{g^{\lambda\sigma}}(e^{i\Gamma_{\rm inv}[g^{\lambda\sigma}]})\,. \tag{6.33}$$

[37]I am assuming that $g^R_{\alpha\beta}$ and $\bar{g}^{\lambda\sigma}$ lie in a closed convex set, so that the conditions of the Schauder fixed point theorem are satisfied. I wish to thank J. and L. Chayes for a conversation about the conditions for the existence of a fixed point.

An alternative way of deriving Eq. (6.32) is to note that when variations of $g^R_{\alpha\beta}$ are included, Eq. (6.20) is modified to read

$$\delta\Gamma=\int d^4x\left[\left[\frac{\delta}{\delta g^R_{\alpha\beta}}W\right]\delta g^R_{\alpha\beta}+\left[\frac{\delta}{\delta J^{\lambda\sigma}}W+\bar{g}_{\lambda\sigma}\right]\delta J^{\lambda\sigma}+\delta\bar{g}_{\lambda\sigma}J^{\lambda\sigma}\right]\,, \tag{6.34}$$

which implies that

$$\frac{\delta}{\delta\bar{g}_{\lambda\sigma}}\Gamma^{\rm inv}[g^{\lambda\sigma}]=\frac{\delta}{\delta g^R_{\lambda\sigma}}W\left[J^{\lambda\sigma}[\bar{g}^{\gamma\delta},\bar{g}_{\alpha\beta}],\bar{g}_{\alpha\beta}\right]+J^{\lambda\sigma}[\bar{g}^{\gamma\delta},\bar{g}_{\alpha\beta}]\,. \tag{6.35}$$

At the solution $\bar{g}_{\alpha\beta}=g^*_{\alpha\beta}$ of the equation

$$J^{\lambda\sigma}[\bar{g}^{\gamma\delta},\bar{g}_{\alpha\beta}]=0\,, \tag{6.36}$$

we learn from Eq. (6.16) that both terms on the right-hand side of Eq. (6.35) are zero, thus reproducing Eq. (6.32).

Having now established the procedure for identifying the background metric and calculating its dynamics, let us restore the matter fields to the analysis. Following the notation of Eqs. (2.1) and (2.38), this is done by making the substitutions

$$\begin{aligned}&d[g_{\mu\nu}]\to d[g_{\mu\nu}]d\{\phi\}\,,\\&S_{\rm grav}[g_{\mu\nu}]\to S_{\rm matter}[\{\phi\},g_{\mu\nu}]+S_{\rm grav}[g_{\mu\nu}]\,,\\&S_{\rm matter}[\{\phi\},g_{\mu\nu}]=\int d^4x\sqrt{-g}\,\mathscr{L}_{\rm matter}[\{\phi\},g_{\mu\nu}]\end{aligned} \tag{6.37}$$

in Eq. (6.10), giving[38]

$$Z=\int d[g_{\mu\nu}]d\{\phi\}\Delta[g^R_{\alpha\beta},g_{\mu\nu}] \times e^{iS_{\rm matter}[\{\phi\},g_{\mu\nu}]+iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]}\,. \tag{6.38}$$

Let us next divide the matter fields $\{\phi\}$ into "light" and "heavy" components[39] as in Sec. II.B, and find the effective action equations governing the dynamics of the light

[38]As in the earlier sections, I do not explicitly indicate the gauge-fixing procedure for the matter gauge fields.

[39]The heavy "matter" fields can include any fields which are not directly observable, including ones which are basically geometric or pregeometric in nature, and auxiliary fields. The only essential requirement for the discussion of Secs. VI.A and VI.B is that the partition function be representable in the form of Eq. (6.38) for some choice of heavy fields $\{\phi^H\}$. The discussion, as given, applies only to the case when the observed matter fields $\{\phi^L\}$ appear as elementary fields in the fundamental action. If, as has been much discussed recently, some of the light fields are effective fields for composites formed from the truly elementary fields, an extended effective action formalism is needed, along the lines discussed by Cornwall, Jackiw, and Tomboulis (1974). For a discussion of the effective action for composites in a nonrelativistic solid-state physics context, see Kleinert (1978).

fields. The most straightforward way of doing this is to introduce external sources $\{J^L\}$ and expectation values $\{\bar{\phi}^L\}$ for the light matter fields $\{\phi^L\}$, as well as an external source $J^{\lambda\sigma}$ and expectation value $\bar{g}^{\lambda\sigma}$ for the metric, and to construct the Legendre-transformed effective action functional $\Gamma[\{\bar{\phi}^L\}, \bar{g}^{\lambda\sigma}, g^R_{\alpha\beta}]$ in analogy with Eqs. (6.15)–(6.21) above. Following Eq. (6.28), the partition function Z of Eq. (6.38) can be expressed in the form

$$Z=\operatorname{ext}_{\bar{g}^{\lambda\sigma},\{\bar{\phi}^L\}}(e^{i\Gamma[\{\bar{\phi}^L\},\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]}),$$

$$\frac{\delta}{\delta g^R_{\alpha\beta}}Z=0, \tag{6.39}$$

and the fixed point argument of Eqs. (6.28)–(6.33) can then be used to show that Eqs. (6.39) are equivalent to

$$Z=\operatorname{ext}_{\bar{g}^{\alpha\beta},\{\bar{\phi}^L\}}(e^{i\Gamma_{\rm inv}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]}), \tag{6.40}$$

with $\Gamma_{\rm inv}$ the general-coordinate invariant effective action

$$\Gamma_{\rm inv}=\Gamma[\{\bar{\phi}^L\},\bar{g}^{\lambda\sigma},\bar{g}_{\alpha\beta}]. \tag{6.41}$$

Equation (6.40) gives an exact description of the dynamics of the light matter-metric system in terms of a classical variational principle

$$\frac{\delta}{\delta \bar{g}^{\alpha\beta}}\Gamma_{\rm inv}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]=\frac{\delta}{\delta\{\bar{\phi}^L\}}\Gamma_{\rm inv}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]=0; \tag{6.42}$$

that is, for an isolated system, the background metric and the light-field expectation values must evolve according to a principle of stationary effective action.

To put Eq. (6.42) in a more familiar form, let us assume the background metric to be slowly varying on the length scale of the heavy fields, so that the curvature dependence of $\Gamma_{\rm inv}$ can be approximated by writing

$$\Gamma_{\rm inv}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]=S_{\rm eff,matter}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]+S_{\rm eff,grav}[\bar{g}_{\alpha\beta}] + \text{small corrections},$$

$$S_{\rm eff,matter}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]=\text{minimal generally covariant extension of } \Gamma_{\rm inv}[\{\bar{\phi}^L\},\eta^{\alpha\beta}]-\Gamma_{\rm inv}[\{0\},\eta^{\alpha\beta}],$$

$$S_{\rm eff,grav}[\bar{g}_{\alpha\beta}]=\Gamma_{\rm inv}[\{0\},\bar{g}^{\alpha\beta}]+O[(\partial_\lambda\bar{g}_{\mu\nu})^4] =\int d^4x\sqrt{-\bar{g}}\frac{1}{16\pi G_{\rm ind}}(\bar{R}-2\Lambda_{\rm ind}), \tag{6.43}$$

with $\bar{R}=R[\bar{g}_{\alpha\beta}]$ the curvature scalar constructed from $\bar{g}_{\alpha\beta}$. As defined in Eq. (6.43), $S_{\rm eff,matter}$ contains terms in $\Gamma_{\rm inv}$ which are $\bar{\phi}^L$ dependent and are of zeroth or first order in space-time derivatives of $\bar{g}_{\alpha\beta}$, while $S_{\rm eff,grav}$ contains terms independent of the matter fields $\bar{\phi}^L$, which are of zeroth through second order in space-time derivatives of $\bar{g}_{\alpha\beta}$. Substituting Eq. (6.43) into Eq. (6.42) gives the classical Einstein equations[40] and the effective classical equations for the matter fields,

$$\frac{1}{8\pi G^{\rm ind}}(\bar{G}^{\mu\nu}+\Lambda_{\rm ind}\bar{g}^{\mu\nu}) =T^{\mu\nu}_{\rm matter} =-\frac{2}{\sqrt{-\bar{g}}}\frac{\delta}{\delta\bar{g}_{\mu\nu}}S_{\rm eff,matter}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}],$$

$$\frac{\delta}{\delta\{\bar{\phi}^L\}}S_{\rm eff,matter}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]=0. \tag{6.44}$$

Of course, making the approximations of Eq. (6.43) is only a matter of convenience in dealing with slowly varying background metrics, and the exact dynamics of $\bar{g}_{\alpha\beta}$ and $\{\bar{\phi}^L\}$, including the effect of higher derivative terms in $\Gamma_{\rm inv}[\{\bar{\phi}^L\},\bar{g}^{\alpha\beta}]$, is always governed by Eq. (6.42).

An alternative way of describing the dynamics of the light fields is to keep them as quantum variables and to introduce, inside the $\{\phi^L\}$ functional integration, an effective action which incorporates the quantum effects of the heavy fields [cf. Eq. (2.15) above]. To do this, we rewrite Eq. (6.38) in the form

$$Z=\int d\{\phi^L\}e^{iW[\{\phi^L\},g^R_{\alpha\beta}]}, \tag{6.45a}$$

$$e^{iW[\{\phi^L\},g^R_{\alpha\beta}]}=\int d[g_{\mu\nu}]d\{\phi^H\}\Delta[g^R_{\alpha\beta},g_{\mu\nu}] \times e^{iS_{\rm matter}[\{\phi\},g_{\mu\nu}]+iS_{\rm grav}[g_{\mu\nu}]+iS_{gf}[g^R_{\alpha\beta},g_{\mu\nu}]}. \tag{6.45b}$$

The dependence of W on $g^R_{\alpha\beta}$ results from the fact that the general covariance of Eq. (6.45b) is broken by the fixed (nonscalar) light fields $\{\phi^L\}$, which act in the same manner as does the source term in Eq. (6.15), and prevent the application of the argument of Eqs. (6.11)–(6.14). Let us now introduce an additional external source $J^{\lambda\sigma}$ for the metric and use it to construct a Legendre-transformed effective action functional for the metric, $\Gamma'[\{\phi^L\},\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]$, as in Eqs. (6.15)–(6.22). [The prime on Γ' is to distinguish it from the functional $\Gamma[\{\bar{\phi}^L\},\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]$ introduced following Eq. (6.38), which was constructed by Legendre transforming with respect to both the metric and the light fields.] This allows us to rewrite Eq. (6.45) in the form

$$Z=\int d\{\phi^L\}\operatorname{ext}_{\bar{g}^{\lambda\sigma}}(e^{i\Gamma'[\{\phi^L\},\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]}), \tag{6.46}$$

[40]Derivations of the Einstein equations similar to that of Eqs. (6.38)–(6.44) have been given by Fradkin and Vilkovisky (1977a, 1977b), by DeWitt (1979), and by Horowitz (1981). Fradkin and Vilkovisky (1977a, 1977b) and DeWitt (1979, 1981) have emphasized that Eq. (6.42) contains corrections to the Einstein equations which are needed for rapidly varying metrics. For discussions of the "out-in" form of the semiclassical gravitational equations, see Kay (1981) and Horowitz (1981).

which gives an exact formulation of the quantum dynamics of the light fields and the background metric, expressed in terms of a general-coordinate noninvariant effective action Γ'. Because the integrand in Eq. (6.46) still depends on $g^R_{\alpha\beta}$, the fixed point argument of Eqs. (6.28)–(6.32) cannot be used to introduce a gauge-invariant effective action inside the light-field functional integration. An alternative way of seeing this is to note that the extremum in Eq. (6.46) makes $\bar{g}^{\lambda\sigma}$ a functional of the integration variables $\{\phi^L\}$, and so the fixed reference metric $g^R_{\alpha\beta}$ cannot be equated to $\bar{g}^{\lambda\sigma}$ inside the functional integration. To proceed further, let us consider the mean-field approximation to Eq. (6.46), obtained by pulling the extremum over $\bar{g}^{\lambda\sigma}$ to the outside of the functional integration (which should be a physically reasonable approximation for the slowly varying components of $\bar{g}^{\lambda\sigma}$),

$$Z_{mf} \approx \text{ext}_{\bar{g}^{\lambda\sigma}} \int d\{\phi^L\} e^{i\Gamma'[\{\phi^L\},\bar{g}^{\lambda\sigma},g^R_{\alpha\beta}]} . \tag{6.47a}$$

Since Z is independent of $g^R_{\alpha\beta}$, Z_{mf} is independent of $g^R_{\alpha\beta}$ to within the accuracy of the mean-field approximation, and so we have

$$\frac{\delta}{\delta g^R_{\alpha\beta}} Z_{mf} \approx 0 . \tag{6.47b}$$

Equations (6.47a,b) have the same structure as Eqs. (6.28a,b) above, and thus within the mean-field approximation we can apply the fixed point argument of Eqs. (6.28)–(6.32), giving

$$Z_{mf} \approx \text{ext}_{\bar{g}^{\alpha\beta}} \int d\{\phi^L\} e^{i\Gamma'_{\text{inv}}[\{\phi^L\},\bar{g}^{\alpha\beta}]} , \tag{6.48}$$

with Γ'_{inv} the general-coordinate invariant effective action

$$\Gamma'_{\text{inv}} = \Gamma'[\{\phi^L\},\bar{g}^{\lambda\sigma},\bar{g}_{\alpha\beta}] . \tag{6.49}$$

Assuming a slowly varying background metric and making an expansion of the primed effective action analogous to that of Eq. (6.43), we can approximate Eq. (6.48) by

$$Z'_{mf} \approx \text{ext}_{\bar{g}^{\alpha\beta}} \int d\{\phi^L\} \times e^{iS'_{\text{eff,matter}}[\{\phi^L\},\bar{g}^{\alpha\beta}]+S'_{\text{eff,grav}}[\bar{g}_{\alpha\beta}]} . \tag{6.50}$$

This gives the field equations for the background metric in the form

$$\frac{1}{8\pi G'_{\text{ind}}}(\bar{G}^{\mu\nu}+\Lambda'_{\text{ind}}\bar{g}^{\mu\nu}) \approx \frac{\int d\{\phi^L\} e^{iS'_{\text{eff,matter}}[\{\phi^L\},\bar{g}_{\alpha\beta}]} T'^{\mu\nu}_{\text{matter}}}{\int d\{\phi^L\} e^{iS'_{\text{eff,matter}}[\{\phi^L\},\bar{g}_{\alpha\beta}]}} = \langle 0^+ | T'^{\mu\nu}_{\text{matter}} | 0^- \rangle ,$$
$$T'^{\mu\nu}_{\text{matter}} = \frac{2}{\sqrt{-\bar{g}}} \frac{\delta}{\delta \bar{g}_{\mu\nu}} S'_{\text{eff,matter}} , \tag{6.51a}$$

with $|0^+\rangle$ and $|0^-\rangle$ the "out" and "in" vacuum states for the observable matter fields. Thus, the background metric formalism, with the mean-field approximation of Eq. (6.47a) and an expansion for slowly varying metrics, gives the "out-in" form[40] of the semiclassical gravitational equations. The quantum field dynamics for the matter fields then follows in the usual fashion from the approximation to the partition function given in Eq. (6.50). Because the induced constants G'_{ind} and Λ'_{ind} do not include the quantum effects of the light matter fields, they are not identical to the constants G_{ind} and Λ_{ind} defined in Eq. (6.43), which do include such effects. However, since one expects

$$G'_{\text{ind}}/G_{\text{ind}} \approx 1 + O[(l_{\text{Planck}}/l_{\text{proton}})^2] , \tag{6.51b}$$

with l_{proton} the proton Compton wavelength, the difference between the primed and unprimed constants is numerically very small.

B. Formulas for G^{-1}_{ind} and Λ_{ind} with a quantized metric

To complete the analysis begun in Sec. VI.A, we must derive expressions for the induced gravitational and cosmological constants in terms of functional integrals over $h_{\mu\nu}$ and the matter fields,[41] and discuss the conditions under which these expressions yield finite answers. Since the gravitational effective action relevant to astronomy and astrophysics is insensitive to the state of motion of the long-wavelength components of the matter fields, it is most convenient to start the derivation of this section from the formula

$$e^{i\Gamma_{\text{inv}}[\bar{g}_{\alpha\beta}]} = \text{ext}_{\{\bar{\phi}^L\}} (e^{i\Gamma_{\text{inv}}[\{\bar{\phi}^L\},\bar{g}_{\alpha\beta}]}) , \tag{6.52}$$

rather than from the functional $\Gamma_{\text{inv}}[\{0\},\bar{g}_{\alpha\beta}]$ of Eq. (6.43). It is also convenient at this point to represent[33] the gravitational compensating determinant $\Delta[\bar{g}_{\alpha\beta},g_{\mu\nu}]$ by an added action density $\sqrt{-g}\,\mathscr{L}_{\text{ghost}}$, and to adopt the convention that a functional argument $h_{\mu\nu}$ implicitly indicates a dependence on the ghost fields and that the integration measure $d[h_{\mu\nu}]$ implicitly includes the ghost integration measure. By substituting the expansion of $\Gamma_{\text{inv}} \approx S_{\text{eff,grav}}$ from Eq. (6.43) into the left-hand side of Eq. (6.52), and noting that the right-hand side of Eq. (6.52) has a functional integral representation obtained by making the substitutions

$$d[h_{\mu\nu}] \rightarrow d[h_{\mu\nu}]d\{\phi\} ,$$
$$S_{\text{grav}} \rightarrow S_{\text{matter}} + S_{\text{grav}} \tag{6.53}$$

[41] In an older terminology, we must compute expressions for the renormalized gravitational and cosmological constants, including radiative corrections arising from virtual metric and matter fluctuations, in terms of the bare parameters appearing in the fundamental Lagrangian.

in Eq. (6.25), we can rewrite Eq. (6.52) in the form[42]

$$e^{i\int d^4x\{\sqrt{-\bar{g}}(1/16\pi G_{\text{ind}})(\bar{R}-2\Lambda_{\text{ind}})+O[(\partial_\lambda \bar{g}_{\mu\nu})^4]\}}$$

$$=\int d[h_{\mu\nu}]d\{\phi\}e^{i\int d^4x\,\bar{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]},$$

$$\bar{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]$$
$$=\sqrt{-g}\,\{\mathscr{L}_{\text{matter}}[\{\phi\},g_{\mu\nu}]+\mathscr{L}_{\text{grav}}[g_{\mu\nu}]$$
$$+\mathscr{L}_{\text{ghost}}[\bar{g}_{\alpha\beta},h_{\mu\nu}]\}$$
$$+\sqrt{-\bar{g}}\,\mathscr{L}_{gf}[\bar{g}_{\alpha\beta},h_{\mu\nu}]-h_{\lambda\sigma}J^{\lambda\sigma}[\bar{g}_{\alpha\beta}]\,,$$

$$g_{\mu\nu}=\bar{g}_{\mu\nu}+h_{\mu\nu}\,. \tag{6.54}$$

Since we now wish to study the effective action at general values of $\bar{g}_{\alpha\beta}$, where it is not stationary, it is essential to retain the source term $J^{\lambda\sigma}[\bar{g}_{\alpha\beta}]$ in $\bar{\mathscr{L}}$. The problem of extracting expressions for G_{ind}^{-1} and Λ_{ind} from Eq. (6.54) has the same formal structure as that set out in Eqs. (5.1) and (5.2) and solved in Sec. V.A. Hence the desired formulas are obtained by making the following substitutions in Eqs. (5.8), (5.14), (5.18), and (5.19),

$$\int d\{\phi\}\to\int d[\;]\equiv\int d[h_{\mu\nu}]d\{\phi\}\,,$$
$$g_{\mu\nu}\to\bar{g}_{\mu\nu}\,,$$
$$S[\{\phi\},\eta_{\mu\nu}]\to\tilde{S}\equiv\int d^4x\,\tilde{\mathscr{L}}[\{\phi\},\eta_{\alpha\beta},h_{\mu\nu}]\,,$$
$$\bar{\mathscr{L}}\to\tilde{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]\,. \tag{6.55}$$

In order to indicate explicitly the appearance of the source current in the following formulas, it is useful to introduce the notation

$$\tilde{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]=\hat{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]-h_{\lambda\sigma}J^{\lambda\sigma}\,,$$
$$\hat{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]=\sqrt{-g}\,[\mathscr{L}_{\text{matter}}+\mathscr{L}_{\text{grav}}+\mathscr{L}_{\text{ghost}}]$$
$$+\sqrt{-\bar{g}}\,\mathscr{L}_{gf}\,,$$
$$J^{\lambda\sigma}[\bar{g}_{\alpha\beta};y]\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}}\equiv\mathscr{J}^{\lambda\sigma}\,,$$
$$\frac{\delta}{\delta\bar{g}_{\mu\nu}(x)}J^{\lambda\sigma}[\bar{g}_{\alpha\beta};y]\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}}\equiv\mathscr{J}^{\lambda\sigma|\mu\nu}(y,x)\,, \tag{6.56}$$

in terms of which

$$\tilde{S}=\int d^4x[\hat{\mathscr{L}}(\{\phi\},\eta_{\alpha\beta},h_{\mu\nu})-h_{\lambda\sigma}\mathscr{J}^{\lambda\sigma}]\,. \tag{6.57}$$

[42]In Eq. (6.54) we have not required the total space-time volume to have a fixed value. Modifications required by a volume constraint and by the presence of boundaries are discussed by Hawking (1979). A volume constraint can be included by adding a Lagrange multiplier term $\kappa_0\sqrt{-g}$ to $\tilde{\mathscr{L}}$, which plays the role of a bare cosmological term and is discussed in more detail in Sec. VI.C below. Space-time boundaries require the addition to the Einstein-Hilbert action of a surface integral over the boundaries. Hasslacher and Mottola (1981) show that when the quantum fluctuations $h_{\mu\nu}$ in Eq. (6.54) are constrained to have zero normal derivative on a boundary, so that the boundary does not fluctuate, a surface term of the expected form automatically appears in the induced gravitational effective action.

The tensors $\mathscr{J}^{\lambda\sigma}$ and $\mathscr{J}^{\lambda\sigma|\mu\nu}$ are implicitly defined by the relations

$$0=\langle h_{\theta\tau}(0)\rangle\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}} \tag{6.58a}$$
$$\propto\int d[\;]e^{i\tilde{S}}h_{\theta\tau}(0)\,,$$

$$0=\frac{\delta}{\delta\bar{g}_{\mu\nu}(x)}\langle h_{\theta\tau}(0)\rangle\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}}$$
$$\propto\int d[\;]e^{i\tilde{S}}V^{\mu\nu}(x)h_{\theta\tau}(0)\,, \tag{6.58b}$$

$$V^{\mu\nu}(x)=2\frac{\delta}{\delta\bar{g}_{\mu\nu}(x)}\int d^4x\,\tilde{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu}]\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}}$$
$$=V_1^{\mu\nu}(x)+V_2^{\mu\nu}(x)\,,$$

$$V_1^{\mu\nu}(x)=2\left[\left[\frac{\partial}{\partial\bar{g}_{\mu\nu}}-\frac{\partial}{\partial x^\lambda}\frac{\partial}{\partial(\partial_\lambda\bar{g}_{\mu\nu})}\right.\right.$$
$$\left.+\frac{\partial}{\partial x^\lambda}\frac{\partial}{\partial x^\sigma}\frac{\partial}{\partial(\partial_\lambda\partial_\sigma\bar{g}_{\mu\nu})}\right]$$
$$\left.\times\hat{\mathscr{L}}[\{\phi\},\bar{g}_{\alpha\beta},h_{\mu\nu};x]\right]\Big|_{\bar{g}_{\alpha\beta}=\eta_{\alpha\beta}}\,,$$

$$V_2^{\mu\nu}(x)=-2\int d^4z\,h_{\lambda\sigma}(z)\mathscr{J}^{\lambda\sigma|\mu\nu}(z,x)\,.$$

After simplifications using Eq. (6.58), the formulas for $\Lambda_{\text{ind}}/G_{\text{ind}}$ and G_{ind}^{-1} take the form

$$-\frac{1}{2\pi}\frac{\Lambda_{\text{ind}}}{G_{\text{ind}}}=\langle V_1(0)\rangle_0\,,$$

$$\frac{1}{16\pi G_{\text{ind}}}=\langle U(0)\rangle_0$$
$$-\frac{i}{96}\int d^4x\,x^2\Big[\langle\mathscr{T}(\bar{V}_1(x)\bar{V}_1(0))\rangle_0$$
$$-\langle\mathscr{T}(V_2(x)V_2(0))\rangle_0\Big]\,,$$

$$\langle A(0)\rangle_0=\frac{\int d[\;]e^{i\tilde{S}}A(0)}{\int d[\;]e^{i\tilde{S}}}\,,$$

$$\langle\mathscr{T}(A(x)B(0))\rangle_0=\frac{\int d[\;]e^{i\tilde{S}}A(x)B(0)}{\int d[\;]e^{i\tilde{S}}}\,,$$

$$\bar{V}_1(x)=V_1(x)-\langle V_1(x)\rangle_0\,,$$
$$V_1(x)=\eta_{\mu\nu}V_1^{\mu\nu}(x)\,,$$
$$V_2(x)=\eta_{\mu\nu}V_2^{\mu\nu}(x)\,,\quad\langle V_2(x)\rangle_0=0\,,$$
$$U(x)=\text{Eq. (5.14) with } g_{\mu\nu}\to\bar{g}_{\mu\nu},\ \bar{\mathscr{L}}\to\hat{\mathscr{L}}\,. \tag{6.59}$$

A second useful formula for $\Lambda_{\text{ind}}/G_{\text{ind}}$ can be obtained by using Eq. (6.35) to calculate the conformal variation of Γ, giving

$$-\frac{1}{2\pi}\frac{\Lambda_{\text{ind}}}{G_{\text{ind}}}=2\eta_{\lambda\sigma}\left[\frac{\delta}{\delta g^R_{\lambda\sigma}}W[\mathscr{J}^{\gamma\delta},\eta_{\alpha\beta}]+\mathscr{J}^{\lambda\sigma}\right]\,. \tag{6.60}$$

Since

$$\frac{\delta}{\delta g^{R}_{\lambda\sigma}} W[\mathscr{J}^{\gamma\delta},\eta_{\alpha\beta}]=0 \quad \text{when } \mathscr{J}^{\lambda\sigma}=0\,, \tag{6.61}$$

we learn from Eq. (6.60) that the condition for the cosmological constant $\Lambda_{\rm ind}$ to vanish is the vanishing of $\mathscr{J}^{\lambda\sigma}=J^{\lambda\sigma}[\eta_{\alpha\beta}]$. This is of course expected, since when Λ_{ind} vanishes, the induced gravitational action $\Gamma_{Inv}[\bar{g}_{\alpha\beta}]$ is stationary at a Minkowski background metric $\eta_{\alpha\beta}$. When $\eta_{\alpha\beta}$ is the stable ground state, the second-order fluctuation operator around $\eta_{\alpha\beta}$ has no negative eigenvalues, and the functional integral formula of Eq. (6.59) is then guaranteed to give a real value for $G_{\rm ind}^{-1}$.

Unlike the situation found in Sec. V.A, where $\langle U(0)\rangle_0$ vanished, the term $\langle U(0)\rangle_0$ in Eq. (6.59) contains nonvanishing contributions quadratic in the fluctuation metric, such as $\langle(\partial h_{\mu\nu}/\partial x^{\lambda})^2\rangle_0$. Hence this term in the formula for $G_{\rm ind}^{-1}$ is qualitatively similar to the relation

$$G_{\rm ind}^{-1}\sim\langle R[h_{\mu\nu}]\rangle_0 \tag{6.62}$$

proposed by Mansouri (1979, 1981), in papers suggesting that Einstein gravitation is generated by dynamical scale-invariance breaking in conformally invariant, order-R^2 gravitational models. However, Eq. (6.62) [which omits the nonlocal V^2 terms of Eq. (6.59)] is not a quantitatively correct expression for $G_{\rm ind}^{-1}$.

C. Conditions for finiteness of $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ and for the vanishing of $\Lambda_{\rm ind}$

Let us turn now to the issue of whether the formulas for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ given in Eq. (6.59) are finite. By construction, the fundamental Lagrangian density $\mathscr{L}_{\rm matter}+\mathscr{L}_{\rm grav}$ contains a complete basis of dimension-four operators formed from the fields which are present, together with a number (say, N) of dimensionless unrenormalized couplings. The dimensional algorithm of Sec. II.C then guarantees that $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ will be calculable in terms of the corresponding N renormalized couplings.[43] If scale invariance remains unbroken, we get $G_{\rm ind}^{-1}=0=\Lambda_{\rm ind}$. If dynamical breaking of scale invariance occurs, we expect one of the N dimensionless couplings to be replaced by a scale mass $\mathscr{M}$, as discussed in Sec. IV.B, and the theory will then yield nonvanishing predictions for $G_{\rm ind}^{-1}$ and $\Lambda_{\rm ind}$ in terms of $\mathscr{M}$ and the remaining $N-1$ dimensionless couplings. The ideal case, of course, would be that in which the fundamental action contains only one dimensionless coupling, so that after dynamical symmetry breaking and dimensional transmutation, no free dimensionless coupling constants remain.

Let us consider next the conditions under which the induced cosmological constant $\Lambda_{\rm ind}$ vanishes, assuming initially that $\mathscr{L}_{\rm matter}+\mathscr{L}_{\rm grav}$ has a unifying symmetry which leaves only a single dimensionless coupling constant, and which requires the vanishing of the bare cosmological constant. Then after dimensional transmutation, $\Lambda_{\rm ind}$ will be calculable in terms of the scale mass $\mathscr{M}$ (which is expected[44] to be in the range $10^{14}-10^{19}$ GeV), but in general $\Lambda_{\rm ind}/\mathscr{M}^2$ will be a number of order unity, in violent contradiction to Eq. (2.23). The only way to save the situation is for the underlying theory to have a "hidden" symmetry which guarantees the vanishing of $\Lambda_{\rm ind}$, as discussed recently by Pagels (1982). The difficulty with implementing this mechanism is that in order for the hidden symmetry to restrict $\Lambda_{\rm ind}$ it must be an unbroken symmetry, and no natural candidate for such a symmetry is known.[45]

An interesting alternative possibility is suggested by recent work in which Ovrut and Wess (1982) use a cosmological constant as a mechanism for breaking supersymmetry. Suppose that the unifying symmetry allows only a single dimensionless coupling constant but does not restrict the value of the bare cosmological constant, so that we can freely add a term $\int d^4x\sqrt{-g}\,\kappa_0$ to the fundamental action. Because κ_0 has dimension four, any polynomial formed from κ_0 and the fields will have dimension greater than or equal to four, and so the added term does not require the introduction of any dimensional renormalization constants with dimension smaller than four. After dynamical symmetry breaking, the theory now has two dimensional parameters, κ_0 and $\mathscr{M}$, or equivalently, $\Lambda_{\rm ind}$ and $\mathscr{M}$. We can then impose as a renormalization condition the requirement that in the absence of real (as opposed to virtual) matter, the Minkowski metric $\eta_{\mu\nu}$ be the stable background metric, which will require[46]

$$\Lambda_{\rm ind}=0=\mathscr{J}^{\lambda\sigma}\,. \tag{6.63}$$

This leaves only one dimensional parameter $\mathscr{M}$, in terms of which all particle masses and Newton's constant are calculable.

In order to implement this alternative mechanism, we must have justifications both for assuming that the bare cosmological constant κ_0 is nonzero, and for imposing

[43] Note, however, that there *is* a dimension-two internal-symmetry scalar operator $\mathscr{O}_2=R[h_{\mu\nu}]$ which transforms as a Lorentz scalar with respect to $\eta_{\mu\nu}$, the limiting value of the background metric $\bar{g}_{\mu\nu}$ appearing in Eq. (6.59). As a consequence, the U and V^2 terms in Eq. (6.59) in general will each be divergent, with the infinities cancelling only in their sum.

[44] This range extends from the so-called "grand unification mass" of particle physics [see Weinberg (1980b) for a review] to the Planck mass.

[45] An unbroken "hidden" symmetry is also required if the unifying symmetry specifies a definite nonzero value for the bare cosmological constant. For a recent survey of quantum gravity with a cosmological constant, see Christensen and Duff (1980).

[46] The fact that stability of the Minkowski metric requires the vanishing of $\Lambda_{\rm ind}$ is noted and used as a renormalization condition in Brout *et al.* (1980).

the renormalization condition that the induced (or renormalized) cosmological constant $\Lambda_{\rm ind}$ vanish. A possible rationale for assuming that κ_0 is nonzero has been given by Hawking (1979), who points out that in order to construct a partition function Z for a fixed total space-time volume one must include a Lagrange multiplier for this volume, and this is formally equivalent to including a bare cosmological term in the fundamental action.[47] A possible rationale for the renormalization condition $\Lambda_{\rm ind}=0$ could be provided by the observation that in a two-parameter theory, the ratio $\Lambda_{\rm ind}/\mathscr{M}^2$ is not constant in nonequilibrium situations. If one could show that nonequilibrium processes in the early universe, such as back-reaction effects from particle production, resulted in the decay of $\Lambda_{\rm ind}$ towards an equilibrium value of zero,[48] then use of the renormalization condition $\Lambda_{\rm ind}=0$ in the equilibrium analysis of Sec. VI.B would be justified.

D. Structure and properties of the fundamental gravitational action

In this final section I will comment very briefly on the structure and on some of the properties of the fundamental gravitational action. I have assumed in the preceding discussion a general order-R^2 gravitational action density of the form

$$\mathscr{L}_{\rm grav}=A_0\mathscr{G}+B_0\mathscr{H}+C_0\mathscr{K}\ , \tag{6.64}$$

with $\mathscr{G},\mathscr{H},\mathscr{K}$ defined in Eq. (2.35). The study of gravitational actions of this type was initiated by Utiyama and DeWitt (1962), and a proof that they lead to a renormalizable perturbation theory has been given by Stelle (1977).[49] When dimensional regularization is employed, all three terms in Eq. (6.64) are in general needed, as shown in detail for the case of a scalar field by Brown (1977) and by Brown and Collins (1980). Even though the action formed from $\mathscr{G}$ is a topological invariant in four dimensions, it is not a topological invariant in 2ω dimensions, and so makes a nontrivial contribution when multiplied by the power series in $(\omega-2)^{-1}$ contained in the coefficient A_0. The only circumstance under which a term $\mathscr{T}(=\mathscr{G},\ \mathscr{H}$, or $\mathscr{K})$ can be omitted from Eq. (6.64) is when the theory with $\mathscr{T}$ deleted has special symmetries, which guarantee that no divergences with the structure of $\mathscr{T}$ are encountered. Thus, for example, a renormalizable theory of matter and gravitation could be formulated without including any order-R^2 terms in the fundamental action, only if $\mathscr{L}_{\rm matter}$ itself had enough symmetry so that no divergences with the structure of $\mathscr{G}$, $\mathscr{H}$, or $\mathscr{K}$ were encountered. Whether such matter actions can be constructed is not presently known. A more realistic possibility for omitting terms from Eq. (6.64) is afforded by the case of classically conformally invariant theories, in which there are hints[50] that the induced $\mathscr{K}$ term may always have a finite coefficient, permitting one to take $C_0=0$ in Eq. (6.64). It is possible to

[47]The value of κ_0 would then presumably be a parameter characterizing the initial quantum fluctuation which led to the birth of the universe.

[48]For a review of gravitational particle production, see Parker (1977), while for an effective action formalism for particle production in the early universe, see Hartle (1977). In an earlier article, Parker (1969), p. 1066, postulated that "the reaction of the particle creation (or annihilation) back on the gravitational field will modify the expansion in such a way as to reduce the creation rate." Since $\Lambda_{\rm ind}>0$ corresponds to positive vacuum energy, a naive extension of this postulate suggests that a state of the early universe with $\Lambda_{\rm ind}>0$ will decay by gravitational particle production to an equilibrium with $\Lambda_{\rm ind}=0$, at which point particle production ceases. Variants of this idea have appeared in models for the creation of the universe through a quantum tunneling event given by Brout *et al.* (1978), Brout *et al.* (1979), Guth (1981), Akatz and Pagels (1982), and Gott (1982). The models of Brout *et al.* and Gott postulate a transition from a particle producing de Sitter phase with $\Lambda_{\rm ind}=0$, $T^{\mu\nu}_{\rm matter}=-\kappa g^{\mu\nu}$, $\kappa\sim l^{-4}_{\rm Planck}$ to a standard equation of state with $P=\frac{1}{3}\rho$ (P=pressure, ρ=density) as a result of back-reaction effects of particle production. When the term $-\kappa g^{\mu\nu}$ is transposed to the $G^{\mu\nu}+\Lambda_{\rm ind}g^{\mu\nu}$ side of the Einstein equations, κ is equivalent to an initially nonvanishing $\Lambda_{\rm ind}/G_{\rm ind}$. Attempts to find an instability associated with $\Lambda_{\rm ind}\neq0$, within the framework of the semiclassical approximation for the coupled matter-metric system [cf. Eq. (6.51) of the text] have not been successful. Abbott and Deser (1982) have shown that the de Sitter solutions obtained when the Einstein equations are solved with $\Lambda_{\rm ind}\neq 0, T^{\mu\nu}=0$ are classically stable against small perturbations. Particle production calculations in de Sitter spaces using the semiclassical formalism have not yielded an unambiguous answer; see Parker (1977), p. 136, and Gibbons (1979), p. 666, for a discussion and references. Hence a dynamical argument to explain why $\Lambda_{\rm ind}\approx 0$ would have to involve nonequilibrium phenomena and/or higher-loop quantum effects which are ignored in the semiclassical approximation.

[49]For further references, see the review of Weinberg (1979).

[50]There are two pieces of evidence that a bare $\mathscr{K}$ term may not be needed in conformally invariant theories, both coming from the study of conformally invariant matter theories on an unquantized background manifold. The first is that apart from a total divergence, the conformal trace anomaly has only $\mathscr{G}$ and $\mathscr{H}$ terms, which implies that the one-loop Lagrangian counterterm contains no divergences proportional to $\mathscr{K}$. [For a succinct discussion and references, see Tsao (1977).] The second is a general formula which Zee (1982b) has recently derived for the coefficient $C_{\rm ind}$ of the induced $\mathscr{K}$ term,

$$C_{\rm ind}=-\frac{1}{13\,824}\int_E d^4x(x^2)^2\Psi(x^2)\ ,$$

in the notation of Eq. (5.44). Since in an asymptotically free gauge theory one has $\Psi\sim(x^2)^{-4}(\log x^2)^{-2}$ for large x^2 [cf. Eq. (5.50)], the integral for $C_{\rm ind}$ is just barely convergent. Zee's formula also shows that $C_{\rm ind}$ is negative definite, and so the theory is free of tachyons; in this connection see also Horowitz (1981) and Yamagishi (1982). Since the gravitational theory of Eq. (6.66) is asymptotically free, it seems a reasonable conjecture that Zee's results will generalize to the case in which the metric is also quantized.

construct renormalizable order-R^2 gravitational theories of greater complexity than Eq. (6.64) by adding new field degrees of freedom in a number of ways (for example, by including torsion[51] or superfields[52]). A prime consideration in searching for the correct gravitational action will almost certainly be that it should unify in a natural way with the fundamental matter action $\mathscr{L}_{\rm matter}$, when that is finally known; this may involve the introduction of "pregeometric" fundamental variables[39,53] which are not directly classifiable as "matter" or "metric."

The momentum space graviton propagator calculated from the fundamental action density of Eq. (6.64) contains a term proportional to

$$\frac{1}{(k^2)^2}=\lim_{m^2\to 0}\frac{1}{m^2}\left[\frac{1}{k^2}-\frac{1}{k^2+m^2}\right]. \tag{6.65}$$

Since the second pole-term in Eq. (6.65) has an unphysical, negative residue, order-R^2 theories do not satisfy unitarity (with positive probabilities) at the tree level. However, unitarity is a statement about the asymptotic scattering states of a field theory and their S-matrix, and hence unlike renormalizability, is a dynamical, rather than a kinematic statement. Thus if radiative corrections play an important role in the dynamics (and they certainly do in theories with dynamical scale-invariance breaking), violations of tree-level unitarity do not necessarily imply violations of unitarity in the full theory. This point was first made a decade ago by Lee and Wick (1969, 1970), who showed that if fields which have negative-residue "ghost" propagators at the tree level become unstable as a result of radiative corrections, then the S matrix for the asymptotic scattering states can obey unitarity with positive probabilities. The relevance of the Lee-Wick mechanism for quantum gravity was first pointed out by Tomboulis (1977) and has since been discussed by a number of authors.[54] As a concrete example [see Hasslacher and Mottola (1981)], let us consider a conformally invariant order-R^2 theory with the fundamental action

$$\begin{aligned}&\mathscr{L}_{\rm grav}=A_0\mathscr{G}+B_0\mathscr{H},\ \ B_0\equiv-\frac{1}{4\xi^2}\,,\\ &\mathscr{G}=\text{Gauss-Bonnet density [Eq. (2.35)]}\,,\\ &\mathscr{H}=C_{\mu\nu\lambda\sigma}C^{\mu\nu\lambda\sigma}\,,\end{aligned} \tag{6.66}$$

with the sign of B_0 chosen to guarantee that the Euclidean continuation of the partition function is represented by a convergent functional integral. (The $\mathscr{G}$ term in the action plays no role in the following discussion and in general does not affect the field equations.) Taking into account the fact that radiative corrections induce an effective Newton's constant, and assuming that $G_{\rm ind}$ has the correct positive sign, a simple calculation shows that the spin-2 part of the full graviton propagator has the form

$$\frac{P^{(2)}_{\mu\nu\alpha\beta}}{k^2[\xi(k^2)^{-2}k^2+m^2(k^2)]}=\frac{P^{(2)}_{\mu\nu\alpha\beta}}{m^2(k^2)}\left[\frac{1}{k^2}-\frac{1}{k^2+\xi(k^2)^2m^2(k^2)}\right]. \tag{6.67}$$

Here $P^{(2)}_{\mu\nu\alpha\beta}$ is a spin-2 projection matrix, $m^2(k^2)$ is the amplitude [analogous to $(d/dk^2)\chi(k^2)$ of Eq. (5.30)] which gives $G_{\rm ind}^{-1}$ in the zero-momentum limit,

$$m^2(0)=\frac{1}{16\pi G_{\rm ind}}\,, \tag{6.68}$$

and $\xi(k^2)$ is the (one-loop) running coupling constant for the action of Eq. (6.66),

$$\xi(k^2)^2=\frac{\xi(\mu^2)^2}{1+\frac{1}{2}b\xi(\mu^2)^2\log(k^2/\mu^2)}\,. \tag{6.69}$$

In the timelike region, where $k^2<0$, both $\xi(k^2)^2$ and $m(k^2)$ have imaginary parts, and consequently the propagator of Eq. (6.67) has two complex conjugate unstable ghost poles rather than a single stable ghost pole. Thus it appears that the Lee-Wick mechanism is applicable to order-R^2 gravitational theories; more detailed checks on this are now needed.

A further property of order-R^2 gravitational theories, which is illustrated by Eqs. (6.67) and (6.69), is that they are asymptotically free. This follows from work of Julve and Tonin (1978), as corrected and extended by Fradkin and Tseytlin (1981) [see also Tomboulis (1980) and Christensen (1982)], showing that $b>0$ in Eq. (6.69) and in the analogous equation for the running coupling constant associated with the $\mathscr{H}$ term in Eq. (6.64). The scale mass $\mathscr{M}$ which characterizes the strong coupling region for the fundamental theory is presumably the Planck mass $m_{\rm Planck}$. At energies much higher than the Planck mass, the theory becomes weakly coupled, and so no singularities are expected.[55] At energies much lower than the Planck mass, the induced gravitational term dominates,

$$\xi(k^2)^{-2}k^2+m^2(k^2)\underset{k^2\to 0}{\to} m^2(0)=\frac{1}{16\pi G_{\rm ind}}\,, \tag{6.70}$$

reflecting the presence of an extra power of k^2 multiplying $\xi(k^2)^{-2}$ in Eqs. (6.67) and (6.70), and giving gravitation the form seen in observational astronomy.

[51] Models with torsion have been discussed by Neville (1980) and by Sezgin and van Nieuwenhuizen (1980), who give further references.

[52] For a discussion of conformal supergravity see Kaku, Townsend, and van Nieuwenhuizen (1978).

[53] For attempts at pregeometric theories of gravitation, see Amati and Veneziano (1981), Terazawa and Akama (1980a, 1980b) and Terazawa (1981a, 1981b).

[54] See Adler (1980b), Hasslacher and Mottola (1981), Tomboulis (1980), and also Salam and Strathdee (1978).

[55] For discussions of singularity avoidance in order-R^2 theories, see Hu (1979), Tomboulis (1980), and Hasslacher and Mottola (1981).

ACKNOWLEDGMENTS

I wish to thank many colleagues in Princeton, Austin, and elsewhere for their helpful questions and comments. In particular, I wish to acknowledge conversations or correspondence with L. S. Brown, J. and L. Chayes, J. C. Collins, R. F. Dashen, A. Duncan, K. Fujikawa, D. J. Gross, B. Hasslacher, B. Holdom, E. Mottola, Y.-J. Ng, B. A. Ovrut, T. Regge, L. Smolin, R. Sorkin, C. Teitelboim, S. Weinberg, J. Wess, H. Yamagishi, A. Zee, and W. Zimmermann. I also wish to thank L. S. Brown, D. G. Boulware, J. C. Collins, A. Duncan, G. T. Horowitz, L. Parker, E. Witten, and A. Zee for their very helpful comments on the first draft of this review. This work was supported by the U. S. Department of Energy under Grant No. DE-ACO2-76ERO2220.

APPENDIX A: DETAILS FOR THE BASIC THEOREMS

1. Arguments excluding dimension-two Lorentz-scalar operators

a. Pure non-Abelian gauge theories in axial and covariant gauges

The necessity for gauge fixing and ghosts requires, in the case of nonAbelian gauge theories, that we give a somewhat more careful argument for the absence of dimension-two Lorentz-scalar and internal symmetry-invariant operators $\mathscr{O}_2$ than would be needed in the Abelian case. Let me give first the argument working in axial gauge

$$A_z^i=0\,. \tag{A1}$$

Since axial gauge is a canonical gauge (Hanson *et al.*, 1976), no ghost fields are present. Hence invariance under the subgroup of the Lorentz group which leaves the z axis invariant and invariance under global internal symmetry transformations restrict a candidate for $\mathscr{O}_2$ to have the form

$$\mathscr{O}_2'=A_x^iA^{ix}+A_y^iA^{iy}+A_t^iA^{it}\,. \tag{A2}$$

Consider now the local gauge transformation

$$\delta A_\mu^i=\partial_\mu\Phi^i-g_0f^{ijk}A_\mu^j\Phi^k\,, \tag{A3}$$

with $\Phi^k=\Phi^k(x,y,t)$ independent of z, so that

$$\delta A_z^i=\partial_z\Phi^i-g_0f^{ijk}A_z^j\Phi^k=0\,. \tag{A4}$$

Under the transformation of Eq. (A3) we have

$$\delta\mathscr{O}_2'=A_x^i\partial_x\Phi^i+A_y^i\partial_y\Phi^i+A_t^i\partial_t\Phi^i\neq 0\,, \tag{A5}$$

and so $\mathscr{O}_2'$ is not invariant under the subclass of local gauge transformations which preserves the $A_z^i=0$ gauge condition. Thus Eq. (A2) is not a physically observable dimension-two Lorentz-scalar operator.

I give next a covariant gauge argument, following the notation of Kugo and Ojima (1979), which uses an inner product ($\cdot$) and an outer product ($\times$) to denote contraction of internal symmetry indices with δ_{ij} and f^{ijk}, respectively. In covariant gauge, we have

$$\mathscr{L}=-\tfrac{1}{4}F_{\mu\nu}\cdot F^{\mu\nu}+\mathscr{L}_{\rm GF}+\mathscr{L}_{\rm FP}\,, \tag{A6}$$

with the gauge-fixing (GF) and Fadde'ev-Popov (FP) Lagrangian terms given by

$$\begin{aligned}\mathscr{L}_{\rm GF}&=-\partial^\mu B\cdot A_\mu+\frac{\alpha_0}{2}B\cdot B\\ &=-\frac{1}{2\alpha_0}(\partial^\mu A_\mu)^2+\frac{\alpha_0}{2}\left[B+\frac{1}{\alpha_0}\partial^\mu A_\mu\right]^2-\partial_\mu(B\cdot A^\mu)\,,\\ \mathscr{L}_{\rm FP}&=-i\partial^\mu\bar{c}\cdot D_\mu c\,,\\ D_\mu&=\partial_\mu-g_0A_\mu\times,\quad c^\dagger=c,\quad \bar{c}^\dagger=\bar{c}\,.\end{aligned} \tag{A7}$$

In Eqs. (A6) and (A7), B is an auxilliary scalar field, α_0 is a gauge parameter, and c is the Fadde'ev-Popov ghost field. The Lagrangian density of Eq. (A6) is invariant under the Becchi-Rouet-Stora (BRS, 1976) transformation

$$\begin{aligned}\delta A_\mu&=\lambda D_\mu c\,,\\ \delta c&=\lambda g_0(c\times c)/2\,,\\ \delta\bar{c}&=i\lambda B\,,\\ \delta B&=0\,,\end{aligned} \tag{A8}$$

with λ an x-independent parameter which anticommutes with c and $\bar{c}$, and all physically observable operators must be similarly invariant. In covariant gauge, Lorentz invariance and invariance under global internal symmetry transformations restrict a candidate for $\mathscr{O}_2$ to have the form (for any constant β_0)

$$\mathscr{O}_2'=A_\mu\cdot A^\mu+\beta_0\bar{c}\cdot c\,. \tag{A9}$$

Under the transformation of Eq. (A8), the change in $\mathscr{O}_2'$ is

$$\delta\mathscr{O}_2'=2A^\mu\cdot\lambda\partial_\mu c+\beta_0[i\lambda Bc+\tfrac{1}{2}\lambda g_0\bar{c}\cdot(c\times c)]\neq 0\,, \tag{A10}$$

and so $\mathscr{O}_2'$ is not a BRS invariant. Hence Eq. (A9) does not give a physically observable dimension-two Lorentz-scalar operator.

We have thus concluded, by working in either axial or covariant gauge, that in a pure non-Abelian gauge theory there are no Lorentz and internal symmetry-invariant operators $\mathscr{O}_2$, and hence no action density terms $\mathscr{O}_2R$ in curved space-time.

b. Massless supersymmetric theories with spin-0 fields

An extension of the above argument excludes Lorentz- and internal symmetry-invariant dimension-two operators $\mathscr{O}_2$ in massless supersymmetric theories with spin-0 fields. Let φ be a massless spin-0 field which has a Majorana spinor supersymmetry partner ψ. Under super-

symmetry transformations, φ transforms[5] as

$$\delta\varphi = i(\bar{\psi}\alpha - \bar{\alpha}\psi) , \tag{A11}$$

with α an x-independent parameter which anticommutes with ψ. Lorentz invariance allows a candidate for $\mathcal{O}_2$ of the form

$$\mathcal{O}'_2 = \varphi^2 , \tag{A12}$$

but under supersymmetry transformations the change in $\mathcal{O}'_2$ is

$$\delta\mathcal{O}'_2 = 2\varphi\delta\varphi \neq 0 . \tag{A13}$$

Hence $\mathcal{O}'_2$ is not an internal symmetry invariant, and an action density term $\mathcal{O}'_2 R$ is excluded. (A dimension-two supersymmetry invariant is readily constructed by adding to $\mathcal{O}'_2$ a fermionic piece proportional to $\bar{\psi}\psi/m_0$, but this requires the introduction of a mass parameter m_0.)

2. Extension to massive regulator schemes

When massive regulators are employed, we learn from the enumeration of Sec. II.C that there are Lagrangian density terms of the form

$$T' = M^4, \quad U' = M^2 R , \tag{A14}$$

with M^4 and M^2 schematically indicating polynomials which are, respectively, quartic and quadratic in the regulator masses. The term T' contributes to the induced cosmological constant $\Lambda_{\rm ind}/G_{\rm ind}$ through the operator $T(0)$ of Eq. (5.19a), while the term U' contributes to the induced gravitational constant $G_{\rm ind}^{-1}$ through the operator $U(0)$ of Eq. (5.19b). The coefficients of T' and U' (which in general depend logarithmically on the regulator masses) are determined by the requirement that $\Lambda_{\rm ind}/G_{\rm ind}$ and $G_{\rm ind}^{-1}$ remain finite as the regulator masses tend to infinity. Consider now the differences

$$\delta(\Lambda_{\rm ind}/G_{\rm ind}) = (\Lambda_{\rm ind}/G_{\rm ind})_{\substack{\rm massive \\ \rm regulator}} - (\Lambda_{\rm ind}/G_{\rm ind})_{\substack{\rm dimensional \\ \rm regularization}} ,$$
$$\delta(G_{\rm ind}^{-1}) = (G_{\rm ind}^{-1})_{\substack{\rm massive \\ \rm regulator}} - (G_{\rm ind}^{-1})_{\substack{\rm dimensional \\ \rm regularization}} , \tag{A15}$$

between the finite induced constants calculated using massive regulators, and the finite values calculated using dimensional regularization. According to the dimensional algorithm, differences such as these between the finite values of connected, one-particle irreducible matrix elements evaluated in two different regularization schemes must be representable as the corresponding matrix elements of a Lagrangian density polynomial $\delta\mathcal{L}$ formed from the bare masses, the bare fields, and $\partial/\partial x^\mu$. The polynomial $\delta\mathcal{L}$ cannot contain the terms T' and U' of Eq. (A14), since any nonzero multiple of these bases is necessarily at least quadratically divergent as the regulator masses tend to infinity. The polynomial $\delta\mathcal{L}$ also cannot contain any field-dependent dimension-four Lagrangian terms which survive in the flat space-time limit, since these would give rise to differences in the flat space-time S matrices calculated in the two regularization schemes. When there are no bare masses and no scalar fields apart from members of massless supermultiplets, no other dimension-four operator is present in curved space-time, and $\delta\mathcal{L}$ then vanishes. We conclude that

$$\delta(\Lambda_{\rm ind}/G_{\rm ind}) = \delta(G_{\rm ind}^{-1}) = 0 ; \tag{A16}$$

that is, under the necessary conditions discussed in Sec. II.D, the renormalized induced gravitational action calculated using massive regulators is unique, and agrees with that calculated by using the method of dimensional regularization.

APPENDIX B: DETAILS FOR THE CALCULATION OF $G_{\rm ind}^{-1}$ IN SU(n) GAUGE THEORY

1. Transformation to one-loop exact renormalization group

In an SU(n) gauge theory, the behavior of physical parameters under changes in the renormalization subtraction point μ is governed [through Eq. (4.18)] by the function $\beta(g)$, which has the power-series expansion

$$\beta(g) = -(\tfrac{1}{2}b_0 g^3 + b_1 g^5 + b_2 g^7 + \cdots) . \tag{B1}$$

Only the first two coefficients $b_{0,1}$ are gauge invariant, and only these coefficients are invariant under coupling constant transformations $g \to g'$ of the form

$$g = g(g') = g' + \sum_{n=1}^{\infty} A_n (g')^{2n+1} , \tag{B2}$$

which are analytic in a neighborhood of $g=0$. 't Hooft (1979) pointed out that the noninvariance of $b_2, \ldots$ under the transformation of Eq. (B2) could be exploited to define a transformation which, in a formal perturbative sense, makes the transformed coefficients $b_2, \ldots$ vanish. Global conditions for the existence of a nonsingular 't Hooft transform were studied by Khuri and McBryan (1979); if singular transformations are not excluded [see Frishman, Horsely, and Wolff (1981) for arguments suggesting the physical relevance of singular coupling constant transformations], then a transformation to a two-loop exact renormalization group can always be made, giving

$$\beta(g) = -(\tfrac{1}{2}b_0 g^3 + b_1 g^5) . \tag{B3}$$

Following Adler (1981), let us now make a further, nonanalytic transformation to a new "reduced" running coupling constant g_R for which a one-loop renormalization group structure is exact. (In the applications of the one-loop exact running coupling constant in Sec. V.D of the text, the subscript R is omitted.) Writing $\bar{\alpha}_R \equiv g_R^2$, $\bar{\alpha} \equiv g^2$, the transformation is simply

$$\frac{1}{\bar{\alpha}_R} = -\frac{1}{2}b_0 \int_{\bar{\alpha}}^{\infty} \frac{d\bar{\alpha}'}{\bar{\beta}(\bar{\alpha}')} ,$$
$$\bar{\beta}(\bar{\alpha}) = g\beta = -(\tfrac{1}{2}b_0\bar{\alpha}^2 + b_1\bar{\alpha}^3) , \tag{B4}$$

which is easily seen to give a nonsingular mapping from the half-line $0<\bar{\alpha}<\infty$ to the half-line $0<\bar{\alpha}_R<\infty$. The renormalization group structure in the new variable $\bar{\alpha}_R$ is determined by $\bar{\beta}_R(\bar{\alpha}_R)$, given by

$$\bar{\beta}_R(\bar{\alpha}_R)=\bar{\beta}(\bar{\alpha})\frac{\partial\bar{\alpha}_R}{\partial\bar{\alpha}}=\bar{\beta}(\bar{\alpha})(-\bar{\alpha}_R^2)\frac{\partial(\bar{\alpha}_R^{-1})}{\partial\bar{\alpha}}=-\frac{1}{2}b_0\bar{\alpha}_R^2 , \tag{B5}$$

and so has exactly a one-loop form.

Explicitly integrating Eq. (B4) gives for the transformation

$$\frac{1}{\bar{\alpha}_R}=\frac{1}{\bar{\alpha}}-a\left[\log\left[\frac{1}{a\bar{\alpha}}\right]+\log(1+a\bar{\alpha})\right] ,$$
$$a=\frac{2b_1}{b_0} , \tag{B6}$$

which for small $a\bar{\alpha}$ can be developed into a series expansion,

$$\frac{1}{\bar{\alpha}_R}=\frac{1}{\bar{\alpha}}-a\log\left[\frac{1}{a\bar{\alpha}}\right]+a\sum_{n=1}^{\infty}\frac{(-a\bar{\alpha})^n}{n} . \tag{B7}$$

Equation (B7) can be inverted to give an expansion for $\bar{\alpha}$ in terms of $\bar{\alpha}_R$ and $\log(a\bar{\alpha}_R)$,

$$\bar{\alpha}=\bar{\alpha}_R(1+\bar{\alpha}_R f) ,$$
$$f=\sum_{k=0}^{\infty}\bar{\alpha}_R^k f_k ,$$
$$f_0=a\log(a\bar{\alpha}_R) ,$$
$$f_1=f_0^2+af_0-a^2 , \ldots . \tag{B8}$$

Because f_0 contains a logarithm, the transformation is nonanalytic at $\bar{\alpha}_R=0$, which is why the coefficient b_1 can be transformed to zero. Substituting Eq. (B8) into a perturbation series which has been brought to 't Hooft's form yields a modified perturbation series in terms of the new running coupling constant g_R, for which the one-loop renormalization group is exact. The modified expansion has the form of a joint power series in $\bar{\alpha}_R$ and $\log(a\bar{\alpha}_R)$ in which, for a physical quantity with leading-order contribution at order $\bar{\alpha}_R^L$, the general term has the form $\bar{\alpha}_R^n[\log(a\bar{\alpha}_R)]^p$, with $n\geq L$ and with $p\leq n-\max(1,L)$.

2. Leading short-distance contribution to $\Psi(t)$

Asymptotic freedom implies that the leading short-distance contribution to $\Psi(t)$ is obtained by doing a lowest-order perturbation theory calculation, with the coupling constant g^2 replaced by the running coupling constant $g^2(t)$. Thus from Eqs. (5.43) and (5.49) we get

$$\Psi(t)=\frac{1}{4}\frac{1}{(-\log\mathcal{M}^2t)^2}[\langle\mathcal{T}(F^2(x)F^2(0))\rangle_{0E}$$
$$-\langle F^2(x)\rangle_{0E}^2] , \tag{B9}$$

with F^2 a shorthand for $F_{\lambda\sigma}^{ri}F^{ri\lambda\sigma}$, and with the subscript $0E$ indicating the Euclidean vacuum expectation. In lowest-order perturbation theory, the square bracket in Eq. (B9) is given by

$$\langle\mathcal{T}(F^2(x)F^2(0))\rangle_{0E}-\langle F^2\rangle_{0E}^2$$
$$=2[\langle\mathcal{T}(F_{\lambda\sigma}^i(x)F_{\mu\nu}^j(0))\rangle_{0E}]^2$$
$$=2[\langle\mathcal{T}(\partial_{[\lambda,}A_{\sigma]}^i(x)\partial_{[\mu,}A_{\nu]}^j(0))\rangle_{0E}]^2 , \tag{B10}$$

with [,] indicating antisymmetrization of indices. Substituting the Euclidean Feynman propagator

$$\langle\mathcal{T}(A_\sigma^i(x)A_\nu^j(y))\rangle_{0E}=\frac{\delta^{ij}\delta_{\sigma\nu}}{(2\pi)^2(x-y)^2} , \tag{B11}$$

and carrying out the differentiations and contractions, an elementary calculation gives

$$[\langle\mathcal{T}(\partial_{[\lambda,}A_{\sigma]}^i(x)\partial_{[\mu,}A_{\nu]}^j(0))\rangle_{0E}]^2$$
$$=\frac{3\times2^7}{(2\pi)^4}(n^2-1)\frac{1}{t^4}$$
$$\Longrightarrow\Psi(t)=\frac{3\times2^6}{(2\pi)^4}(n^2-1)\frac{1}{t^4(-\log\mathcal{M}^2t)^2} , \tag{B12}$$

yielding the value of C_Ψ given in Eq. (5.52) of the text.

3. Dimensional continuation evaluation of comparison integrals

I give here two evaluations of the integral of Eq. (5.55) by dimensional continuation. In the first calculation, only the power of u in the integrand is dimensionally continued, while the logarithms are kept in dimension four. In the second calculation [restricted for simplicity to the leading term in $\Theta(u)$] both the power of u and the logarithms are dimensionally continued, corresponding to use of the 2ω-dimensional vacuum expectation in Eq. (5.22). The two calculations give the same answer, as expected where a finite radiative correction is evaluated by different regularization methods. In the context of the second calculation, we can compare the analyticity properties in ω of the dimensional continuation of a finite sum of Feynman diagrams, with the analyticity properties of the infinite sum of Feynman diagrams contained in the running coupling constant factor $g^4(t)$.

In 2ω dimensions, the factor $d^{2\omega}x$ in Eq. (5.22) is proportional to $dt\,t^{\omega-1}$, and since $\bar{T}(x)$ has canonical dimension 2ω, the leading power behavior of the vacuum expectation $\langle\mathcal{T}(\bar{T}(x)\bar{T}(0))\rangle_0^\omega$ is $t^{-2\omega}$. Hence when the logarithmic sum $\Theta(u)/(\log u)^2$ is kept in four dimensions (and when a normalization factor of π^ω/π^2 is omitted), the continuation of the integral of Eq. (5.55) is

$$\int_0^{u_0}(du\,u^{\omega-1})u\left[u^{-2\omega}\frac{\Theta(u)}{(\log u)^2}\right]=\int_0^{u_0}du\,u^{-\omega}\frac{\Theta(u)}{(\log u)^2} , \tag{B13}$$

and is convergent at $u=0$ when $\mathrm{Re}\,\omega<1$. In order to put Eq. (B13) in a form where it can be analytically con-

tinued to $\omega=2$, let us first make the change of variable $u=e^{-v}$, giving

$$\int_{\log u_0^{-1}}^{\infty} dv \frac{e^{(\omega-1)v}}{v^2}\Theta(e^{-v}) , \tag{B14}$$

with the contour of integration running along the positive real axis. When $\mathrm{Re}\omega<1$ and $\mathrm{Im}\omega>0$, the integration contour can be deformed to the contour C of Fig. 5, while when $\mathrm{Re}\omega<1$ and $\mathrm{Im}\omega<0$, the contour can be deformed to a contour C^*, obtained by reflecting C in the real axis. Once the contour has been deformed to C or C^*, the integral of Eq. (B14) converges for any value of $\mathrm{Re}\omega$, and we can continue $\mathrm{Re}\omega$ to 2. Since Hermiticity of a quantum field theory requires that the regularization prescription be manifestly real (contour prescriptions can enter only through Feynman propagators), the limit as $\omega\to2$ must be defined as the average of dimensional continuations to $\omega=2+i\epsilon$ and to $\omega=2-i\epsilon$. That is, we must average the evaluations of Eq. (B14) with $\omega=2$ on the contours C and C^*, or equivalently, take the real part of the evaluation on the contour C alone, yielding the formula given in Eq. (5.56) of the text. The inequivalence of the evaluations on C and C^* implies that the analytic continuation of Eq. (B14) to $\mathrm{Re}\omega>1$ has a branch cut running along the positive real axis from $\omega=1$ (space-time dimension two) to ∞.

To study the effect of dimensionally continuing the logarithmic terms in Eq. (5.55), we note that a momentum space factor $\log k^2$ continues into

$$\frac{(k^2)^{\omega-2}}{\omega-2}+\text{counterterm}=\frac{(k^2)^{\omega-2}-1}{\omega-2} , \tag{B15}$$

and corresponding to this, a coordinate space factor $\log(-x^2)=\log t$ continues into

$$\frac{(x^2)^{2-\omega}}{2-\omega}+\text{counterterm}=\frac{(x^2)^{2-\omega}-1}{2-\omega} . \tag{B16}$$

Hence let us consider the integral

$$I(\omega,\gamma)=\int_0^{u_0} du\, u^{-\omega}\frac{1}{\left[\dfrac{u^\gamma-1}{\gamma}\right]^2} , \tag{B17}$$

which when $\gamma=2-\omega$ describes the continuation of the leading logarithmic term in $\Theta(u)$, and when $\gamma=0$ reduces to the integral, studied above, in which the logarithmic factor is not continued,

$$I(\omega,0)=\int_0^{u_0} du\, u^{-\omega}\frac{1}{(\log u)^2} . \tag{B18}$$

To study the ω-plane analyticity of Eq. (B17), we expand the factor $(1-u^\gamma)^{-2}$ into a power series in u^γ (since $u_0<1$, this is permitted for $\gamma>0$), and then do the u integrations assuming $\mathrm{Re}\omega<1$, giving

$$\begin{aligned} I(\omega,\gamma)&=\gamma^2\int_0^{u_0} du\, u^{-\omega}\sum_{n=0}^{\infty}(n+1)u^{n\gamma} \\ &=\gamma^2\sum_{n=0}^{\infty}(n+1)\frac{u_0^{n\gamma+1-\omega}}{n\gamma+1-\omega} . \end{aligned} \tag{B19}$$

When γ is regarded as a parameter independent of ω, Eq. (B19) shows that $I(\omega,\gamma)$ is a meromorphic function of ω, with poles at $\omega=1+n\gamma$, $n=0,1,\ldots$. In the limit as $\gamma\to0$ for fixed ω, these poles coalesce into a branch cut running from $\omega=1$ to $\omega=+\infty$, which is just the analyticity structure of $I(\omega,0)$ which was inferred from the discussion following Eq. (B14) above. When the value $\gamma=2-\omega$, corresponding to continuation of the logarithm, is substituted into Eq. (B19), we get

$$\begin{aligned} I(\omega,2-\omega)&=(2-\omega)^2\sum_{n=0}^{\infty}(n+1)\frac{u_0^{2n+1-\omega(n+1)}}{2n+1-\omega(n+1)} \\ &=(2-\omega)u_0^{1-\omega}\,{}_2F_1\left[2,\frac{1-\omega}{2-\omega};\frac{1-\omega}{2-\omega}+1;u_0^{2-\omega}\right] , \end{aligned} \tag{B20}$$

with the hypergeometric function ${}_2F_1$ $(a=2,b;$ $c=b+1;z)$ defined by

$${}_2F_1(2,b;b+1;z)=\sum_{n=0}^{\infty}\frac{(n+1)z^n}{b+n} . \tag{B21}$$

The singularities of Eq. (B21) are poles at $b_n=-n$, corresponding to poles in ω at ω_n given by

$$1\le\omega_n=\frac{2n+1}{n+1}<2 , \tag{B22}$$

and a cut along the real z axis from $z=1$ to $z=\infty$, corresponding to a cut along the real ω axis from $\omega=2$ to $\omega=\infty$. Hence there is an infinite accumulation of poles on the real axis to the left of $\omega=2$, and a branch cut on the real axis to the right of $\omega=2$. As a result, the limit $\omega\to2$ cannot be taken along the real axis, and instead must be defined as the average of limits from above and below the real axis, giving the real part prescription of Eq. (5.56). The fact that $\omega=2$ is a branch point is a direct result of the fact that $I(\omega,2-\omega)$ is the sum of an infinite number of Feynman diagrams. If the sum in Eq. (B20) is truncated at $n=N$, corresponding to retaining only contributions to the running coupling constant through N-loop order, one gets a meromorphic function of ω which is regular at $\omega=2$. This is the result expected from the discussion of the dimensional continuation of individual Feynman diagrams given in Sec. III.

We must still show that, as $\omega\to2$ in a real part or principal value sense, $I(\omega,2-\omega)$ approaches the leading term of Eq. (5.56) of the text. To do this, let us again make the change of variable $u=e^{-v}$, giving

$$I(\omega,2-\omega)=\int_{\log u_0^{-1}}^{\infty} dv \frac{e^{(\omega-1)v}}{v^2}\frac{1}{\left[\dfrac{1-e^{(\omega-2)v}}{(2-\omega)v}\right]^2} . \tag{B23}$$

For $\mathrm{Re}\omega<1$, $\mathrm{Im}\omega>0$, and v in the first quadrant, we have

$$\begin{aligned} &\mathrm{Re}[(\omega-1)v]=\mathrm{Re}(\omega-1)\mathrm{Re}v-\mathrm{Im}(\omega-1)\mathrm{Im}v<0 , \\ &\mathrm{Re}[(\omega-2)v]=\mathrm{Re}[(\omega-1)v]-\mathrm{Re}v<0 , \end{aligned} \tag{B24}$$

and so the contour of integration can be deformed to C without encountering poles coming from vanishing of the denominator in Eq. (B23). We can then set $\omega=2+i\epsilon$, giving

$$\mathrm{Re}[I(2+i\epsilon,-i\epsilon)]=\mathrm{Re}\left[\int_{\log u_0^{-1}}^{i\infty}\frac{dv}{v^2}e^{v}F(v,\epsilon)\right], \tag{B25}$$

with $F(v,\epsilon)$ given by

$$F(v,\epsilon)=\frac{e^{i\epsilon v}}{\left[\frac{1-e^{i\epsilon v}}{i\epsilon v}\right]^2},$$

$$F(v,\epsilon)=1+O(\epsilon^2|v|^2),\quad |v|\ll\epsilon^{-1},$$

$$F(v,\epsilon)=\frac{1}{\left[\frac{\sinh\frac{1}{2}|\epsilon v|}{\frac{1}{2}|\epsilon v|}\right]}\le 1,\quad v \text{ imaginary}. \tag{B26}$$

Since the integral of Eq. (B25) is absolutely and uniformly convergent for all $\epsilon\geq 0$, we can take the limit as $\epsilon\rightarrow 0$ inside the integral, giving

$$I(2,0)=\lim_{\epsilon\rightarrow 0}\mathrm{Re}[I(2+i\epsilon,-i\epsilon)]$$
$$=\mathrm{Re}\left[\int_{\log u_0^{-1}}^{i\infty}\frac{dv}{v^2}e^{v}\right]. \tag{B27}$$

The result of this rather tedious analysis thus reproduces the leading, $\Theta=1$, term of Eq. (5.56).

REFERENCES

Abbott, L. F., 1981, Nucl. Phys. B **185**, 189.
Abbott, L. F., and S. Deser, 1982, Nucl. Phys. B **195**, 76.
Abers, E. S., and B. W. Lee, 1973, Phys. Rep. **9**, 1.
Adler, S. L., 1976, Phys. Rev. D **14**, 379.
Adler, S. L., 1980a, Phys. Rev. Lett. **44**, 1567.
Adler, S. L., 1980b, Phys. Lett. B **95**, 241.
Adler, S. L., 1980c, "Induced gravitation," to appear in *The High Energy Limit* (the 1980 Erice lectures), edited by A. Zichichi (Plenum, New York).
Adler, S. L., 1981, Phys. Rev. D **23**, 2905, Appendix B.
Adler, S. L., J. C. Collins, and A. Duncan, 1977, Phys. Rev. D **15**, 1712.
Adler, S. L., J. Lieberman, Y. J. Ng, and H. S. Tsao, 1976, Phys. Rev. D **14**, 359.
Akama, K., Y. Chikashige, T. Matsuki, and H. Terazawa, 1978, Prog. Theor. Phys. **60**, 868.
Akatz, D., and H. Pagels, 1982, Phys. Rev. D **25**, 2065.
Amati, D., and G. Veneziano, 1981, Phys. Lett. B **105**, 358.
Ambjörn, J., and P. Olesen, 1980, Nucl. Phys. B **170**, 60.
Banks, T., R. Horsley, H. R. Rubinstein, and U. Wolff, 1981, Nucl. Phys. B **190**, 692.
Bardeen, J., L. N. Cooper, and J. R. Schrieffer, 1957, Phys. Rev. **106**, 162.
Batalin, I. A., and E. S. Fradkin, 1979, Phys. Lett. B **86**, 263.
Batalin, I. A., S. G. Matinyan, and G. K. Savvidi, 1977, Yad. Fiz. **26**, 407 [Sov. J. Nucl. Phys. **26**, 214].
Becchi, C., A. Rouet, and R. Stora, 1976, Ann. Phys. (N.Y.) **98**, 287.
Berg, B., 1981, Phys. Lett. B **97**, 401.
Bernard, C., 1979, Phys. Rev. D **19**, 3013.
Bernstein, J., 1974, Rev. Mod. Phys. **46**, 7.
Bhanot, G., and C. Rebbi, 1981, Nucl. Phys. B **180**, 469.
Binder, K., 1976, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York), Vol. 56, pp. 1–101.
Bjorken, J. D., and S. D. Drell, 1965, *Relativistic Quantum Fields* (McGraw-Hill, New York).
Boulware, D. G., 1981, Phys. Rev. D **23**, 389.
Boulware, D. G., 1982, work in progress.
Brout, R., F. Englert, J-M. Frère, E. Gunzig, P. Nardone, and C. Truffin, 1980, Nucl. Phys. B **170**, 228.
Brout, R., F. Englert, and E. Gunzig, 1978, Ann. Phys. (Paris) **115**, 78.
Brout, R., F. Englert, and P. Spindel, 1979, Phys. Rev. Lett. **43**, 417.
Brown, L. S., 1977, Phys. Rev. D **15**, 1469.
Brown, L. S., 1980, Ann. Phys. (N.Y.) **126**, 135.
Brown, L. S., and J. C. Collins, 1980, Ann. Phys. (N.Y.) **130**, 215.
Brown, L. S., and A. Zee, 1982, work in progress.
Callan, C. G., 1970, Phys. Rev. D **2**, 1541.
Chanowitz, M. S., and J. Ellis, 1972, Phys. Lett. B **40**, 397.
Chanowitz, M. S., and J. Ellis, 1973, Phys. Rev. D **7**, 2490.
Christensen, S. M., 1982, to appear in *Quantum Structure of Space and Time*, edited by M. J. Duff and C. J. Isham (Cambridge University, Cambridge).
Christensen, S., and M. Duff, 1980, Nucl. Phys. B **170**, 480.
Chudnovsky, E. M., 1978, Teor. Mat. Fiz. **35**, 398 [Theor. Math. Phys. (USSR) **35**, 538].
Coleman, S., 1979, in *The Whys of Subnuclear Physics* (the 1977 Erice lectures), edited by A. Zichichi (Plenum, New York) pp. 805–941.
Coleman, S., and R. Jackiw, 1971, Ann. Phys. (N.Y.) **67**, 552.
Coleman, S., and E. Weinberg, 1973, Phys. Rev. D **7**, 1888.
Collins, J. C. (unpublished).
Collins, J. C., 1975, *Dimensional Regularization and Renormalization of Quantum Field Theory*, Ph.D. thesis (King's College, Cambridge).
Collins, J. C., A. Duncan, and S. D. Joglekar, 1977, Phys. Rev. D **16**, 438.
Cornwall, J. M., R. Jackiw, and E. Tomboulis, 1974, Phys. Rev. D **10**, 2428.
Cornwall, J. M., and R. E. Norton, 1973, Phys. Rev. D **8**, 3338.
Creutz, M., 1978, Rev. Mod. Phys. **50**, 561.
Creutz, M., 1980, Phys. Rev. D **21**, 2308.
Creutz, M., L. Jacobs, and C. Rebbi, 1979, Phys. Rev. D **20**, 1915.
Crewther, R. J., 1972, Phys. Rev. Lett. **28**, 1421.
DeWitt, B. S., 1965, *Dynamical Theory of Groups and Fields*, (Gordon and Breach, New York).
DeWitt, B. S., 1979, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge), p. 680.
DeWitt, B. S., 1981, to appear in *Quantum Gravity II*, edited by C. J. Isham, R. Penrose and D. W. Sciama (Oxford University, Oxford).
Di Giacomo, A., and G. Paffuti, 1982, Phys. Lett. B **108**, 327.
Duff, M., 1981, to appear in *Quantum Gravity II*, edited by C.

J. Isham, R. Penrose and D. W. Sciama (Oxford University, Oxford).
Englert, F., and R. Brout, 1964, Phys. Rev. Lett. 13, 321.
Englert, F., C. Truffin, and R. Gastmans, 1976, Nucl. Phys. B 117, 407.
Erdélyi, A., ed., 1954, *Tables of Integral Transforms* (McGraw-Hill, New York).
Fadde'ev, L. D., and V. N. Popov, 1967, Phys. Lett. B **25**, 29.
Fadde'ev, L. D., and A. A. Slavnov, 1980, *Gauge Fields* (Benjamin/Cummings, Reading, Mass.).
Fayet, P., and S. Ferrara, 1977, Phys. Rep. **32**, 249.
Fetter, A. L., and J. D. Walecka, 1971, *Quantum Theory of Many Particle Systems* (McGraw-Hill, New York).
Fradkin, E. S., and A. A. Tseytlin, 1981, Phys. Lett. B **104**, 377.
Fradkin, E. S., and G. A. Vilkovisky, 1975, Phys. Lett. B 55, 224.
Fradkin, E. S., and G. A. Vilkovisky, 1976, "On Renormalization of Quantum Field Theory in Curved Spacetime," Berne preprint.
Fradkin, E. S., and G. A. Vilkovisky, 1977a, in *Proceedings of the 18th International Conference on High Energy Physics, Tbilisi, U.S.S.R., 1976* (Dubna), Vol. 2, Sec. C, p. 28.
Fradkin, E. S., and G. A. Vilkovisky, 1977b, Lett. Nuovo Cimento 19, 47.
Frishman, Y., R. Horsley, and U. Wolff, 1981, Nucl. Phys. B **183**, 509.
Fritzsch, H., and P. Minkowski, 1981, Phys. Rep. 73, 87.
Fujii, Y., 1974, Phys. Rev. D 9, 874.
Fujikawa, K., 1981, Phys. Rev. D 23, 2262.
Fukuda, R., and Y. Kazama, 1980, Phys. Rev. Lett. **45**, 1142.
Gibbons, G., 1979, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge), p. 639.
Gott, J. R., III, 1982, Nature **295**, 304.
Gottlieb, S., and J. T. Donohue, 1979, Phys. Rev. D **20**, 3378.
Gross, D. J., and A. Neveu, 1974, Phys. Rev. D 10, 3235.
Gross, D. J., and F. Wilczek, 1973, Phys. Rev. Lett. **30**, 1343.
Gross, D. J., and F. Wilczek, 1974, Phys. Rev. D 9, 980.
Guth, A. H., 1981, Phys. Rev. D 23, 347.
Hanson, A., T. Regge, and C. Teitelboim, 1976, *Constrained Hamiltonian Systems* (Acad. Naz. dei Lincei, Rome), Vol. 373.
Hartle, J. B., 1977, Phys. Rev. Lett. 39, 1373.
Hasslacher, B., and E. Mottola, 1980, Phys. Lett. B **95**, 237.
Hasslacher, B., and E. Mottola, 1981, Phys. Lett. B 99, 221.
Hawking, S. W., 1979, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge), p. 746.
Hawking, S. W., and W. Israel, 1979, *General Relativity* (Cambridge University, Cambridge).
Heisenberg, W., and H. Euler, 1936, Z. Phys. 98, 714.
't Hooft, G. (unpublished).
't Hooft, G., 1975, in *Functional and Probabilistic Methods in Quantum Field Theory*, edited by Bernard Jancewicz, Acta Universitatis Wratislavensis No. 368, XIIth Winter School of Theoretical Physics in Karpacz, February 17-March 2, 1975, Vol. I, p. 345.
't Hooft, G., 1979, in *The Whys of Subnuclear Physics* (the 1977 Erice lectures), edited by A. Zichichi, (Plenum, New York), pp. 943–971.
't Hooft, G., and M. Veltman, 1972, private communication.
Horowitz, G. T., 1981, in *Quantum Gravity II*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford University, Oxford).
Hu, B. L., 1979, "Quantum field theories and relativistic cosmology," to be published in *Recent Developments in General Relativity*, edited by R. Ruffini, Proceedings of the Second Marcel Grossmann Meeting, Trieste, Italy.
Ichinose, S., and M. Omote, 1982, Nucl. Phys. B (in press).
Jackiw, R., and K. Johnson, 1973, Phys. Rev. D 8, 2386.
Johnson, K., M. Baker, and R. Willey, 1964, Phys. Rev. **136**, 1111B.
Julve, J., and M. Tonin, 1978, Nuovo Cimento B **46**, 137.
Kaku, M., P. K. Townsend, and P. van Nieuwenhuizen, 1978, Phys. Rev. D 17, 3179.
Kallosh, R. E., 1978, Nucl. Phys. B **141**, 141.
Kataev, A. L., N. V. Krasnikov, and A. A. Pivovarov, 1982, Nucl. Phys. B (in press).
Kay, B. S., 1981, Phys. Lett. B **101**, 241.
Kazama, Y., 1980, in Proceedings of the XXth International Conference on High Energy Physics, edited by L. Durand and L. G. Pondrum, A.I.P. Conference Proceedings No. 68, Particles and Fields Subseries No. 22, p. 1013.
Khuri, N. N., and O. A. McBryan, 1979, Phys. Rev. D **20**, 881.
Kim, J. E., P. Langacker, M. Levine, and H. H. Williams, 1981, Rev. Mod. Phys. 53, 211.
Klein, O., 1974, Phys. Scr. 9, 69.
Kleinert, H., 1978, Fortschr. Phys. **26**, 565.
Kogut, J., and L. Susskind, 1975, Phys. Rev. D 11, 395.
Kripfganz, J., 1981, Phys. Lett. B 101, 169.
Kugo, T., and I. Ojima, 1979, Prog. Theor. Phys. Suppl. 66.
Lane, K., 1974a, Phys. Rev. D 10, 1353.
Lane, K., 1974b, Phys. Rev. D 10, 2605.
Lee, T. D., and G. C. Wick, 1969, Nucl. Phys. B 9, 209; B 10, 1.
Lee, T. D., and G. C. Wick, 1970, Phys. Rev. D 2, 1033.
Leibbrandt, G., 1975, Rev. Mod. Phys. **47**, 849.
Linde, A. D., 1979, Zh. Eksp. Teor. Fiz. Pis'ma Red 30, 479 [JETP Lett. 30, 447].
Linde, A. D., 1980, Phys. Lett. B 93, 394.
Mansouri, F., 1979, Phys. Rev. Lett. **42**, 1021.
Mansouri, F., 1981, "Conformal Gravity, Conformal Algebra in Minkowski Space, and Dynamical Breakdown of Local Scale Invariance," Yale preprint YTP81-12.
Marciano, W., and H. Pagels, 1978, Phys. Rep. 36, 137.
Matsuki, T., 1978, Prog. Theor. Phys. 59, 235.
Milton, K. A., 1981, Phys. Lett. B **104**, 49.
Minkowski, P., 1977, Phys. Lett. B 71, 419.
Misner, C. W., K. S. Thorne, and J. A. Wheeler, 1970, *Gravitation* (Freeman, San Francisco).
Nambu, Y., and G. Jona-Lasinio, 1961, Phys. Rev. **122**, 345.
Neville, D. E., 1980, Phys. Rev. D 21, 867.
Nieh, H. T., 1982, Phys. Lett. A 88, 388.
Nielsen, N. K., 1977, Nucl. Phys. B **120**, 212.
Nielsen, N. K., 1978, Nucl. Phys. B **140**, 499.
Ovrut, B. A., 1981, "Axial vector ward identities and dimensional regularization" (submitted to Nucl. Phys.).
Ovrut, B. A., and H. J. Schnitzer, 1980, Phys. Rev. D 21, 3369.
Ovrut, B. A., and H. J. Schnitzer, 1981, Nucl. Phys. B 189, 509.
Ovrut, B. A., and J. Wess, 1982, Phys. Lett. B (in press).
Pagels, H., 1982, "Vacuum energy and supergravity," to appear in *Orbis Scientiae 1981*, edited by A. Perlmutter and L. F. Scott, (Plenum, New York).
Pagels, H., and E. Tomboulis, 1978, Nucl. Phys. B **143**, 485.

Parker, L., 1969, Phys. Rev. 183, 1057.
Parker, L., 1977, in *Asymptotic Structure of Space-Time,* edited by F. P. Esposito and L. Witten, (Plenum, New York), p. 107.
Politzer, H. D., 1973, Phys. Rev. Lett. 30, 1346.
Sakharov, A. D., 1967, Dok. Akad. Nauk. SSSR 177, 70 [Sov. Phys. Dokl. 12, 1040 (1968)].
Sakharov, A., 1975, Teor. Mat. Fiz. 23, 178 [Theor. Math. Phys. (USSR) 23, 435].
Salam, A., 1968, in *Elementary Particle Theory,* edited by N. Svartholm (Almqvist and Wiksell, Stockholm), p. 367.
Salam, A., and J. Strathdee, 1978, Phys. Rev. D 18, 4480.
Savvidy, G. K., 1977, Phys. Lett. B 71, 133.
Schrieffer, J. R., 1964, *Theory of Superconductivity* (Benjamin, New York).
Schwinger, J., 1951, Phys. Rev. 82, 664.
Sezgin, E., and P. van Nieuwenhuizen, 1980, Phys. Rev. D 21, 3269.
Simon, B., 1981, *Large Orders and Summability of Eigenvalue Perturbation Theory: A Mathematical Overview,* to appear in Int. J. Quantum Chem. (Proceedings of the 1981 Sanibel Workshop).
Smolin, L., 1979, Nucl. Phys. B 160, 253.
Stelle, K. S., 1977, Phys. Rev. D 16, 953.
Stevenson, P. M., 1981, Ann. Phys. (N.Y.) 132, 383.
Susskind, L., 1979, Phys. Rev. D 20, 2619.
Symanzik, K., 1970, Commun. Math. Phys. 18, 227.
Terazawa, H., 1981a, Phys. Lett. B 101, 43.
Terazawa, H., 1981b, "Pregeometry," to appear in the *Proceedings of the Second Seminar "Quantum Gravity," Academy of Sciences of the USSR, Moscow, October 13–15, 1981.*
Terazawa, H., and K. Akama, 1980a, Phys. Lett. B 96, 276.
Terazawa, H., and K. Akama, 1980b, Phys. Lett. B 97, 81.
Terazawa, H., Y. Chikashige, K. Akama, and T. Matsuki, 1977a, Phys. Rev. D 15, 1181.
Terazawa, H., Y. Chikashige, K. Akama, and T. Matsuki, 1977b, J. Phys. Soc. (Japan) 43, 5.
Tomboulis, E., 1977, Phys. Lett. B 70, 361.
Tomboulis, E., 1980, Phys. Lett. B 97, 77.
Tsao, H.-S., 1977, Phys. Lett. B 68, 79.
Utiyama, R., and B. S. DeWitt, 1962, J. Math. Phys. 3, 608.
Vainstein, A. I., V. I. Zakharov, and M. A. Shifman, 1978, Zh. Eksp. Teor. Fiz. Pis'ma Red 27, 60 [JETP Lett. 27, 55].
Van Nieuwenhuizen, P., 1981, Phys. Rep. 68, 189.
Weinberg, S., 1957, Phys. Rev. 106, 1301.
Weinberg, S., 1967, Phys. Rev. Lett. 19, 1264.
Weinberg, S., 1972, Phys. Rev. Lett. 29, 388.
Weinberg, S., 1976, Phys. Rev. D. 13, 974.
Weinberg, S., 1979, in *General Relativity,* edited by S. W. Hawking, and W. Israel (Cambridge University, Cambridge), p. 790.
Weinberg, S., 1980a, Phys. Lett. B 91, 51.
Weinberg, S., 1980b, Rev. Mod. Phys. 52, 515.
Wilson, K., 1968, Phys. Rev. 179, 1499.
Wilson, K. G., 1973, Phys. Rev. D 7, 2911.
Wilson, K., 1974, Phys. Rev. D 10, 2445.
Witten, E., and S. Weinberg, 1980, Phys. Lett. B 96, 59.
Yamagishi, H., 1982, Phys. Lett. B (in press).
Zee, A., 1979, Phys. Rev. Lett. 42, 417.
Zee, A., 1981, Phys. Rev. D 23, 858.
Zee, A., 1982a, Phys. Rev. Lett. 48, 295.
Zee, A., 1982b, Phys. Lett. B 109, 183.
Zee, A., 1982c, *Gravity as a Dynamical Consequence of the Strong, Weak and Electromagnetic Interactions* (the 1981 Erice Lectures), edited by A. Zichichi (to be published by Plenum, New York).
Zel'dovich, Ya. B., 1967, Zh. Eksp. Teor. Fiz. Pis'ma Red 6, 883 [JETP Lett 6, 316].
Zimmermann, W., 1970, in *Lectures on Elementary Particles and Quantum Field Theory* (the 1970 Brandeis lectures), edited by S. Deser, M. Grisaru, and H. Pendleton (MIT, Cambridge, Mass.), pp. 395–589.

Erratum: Einstein gravity as a symmetry-breaking effect in quantum field theory [Rev. Mod. Phys. 54, 729 (1982)]

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540

The following clarifications should help in reading Sec. VI:

(1) Equation (6.11) is obtained by substituting Eq. (6.9), with $\theta'=\theta$, into Eq. (6.7). The step from Eq. (6.13) to (6.14) then makes use of Eq. (6.11), with θ replaced by θ^{-1} and with $g'^{R}_{\alpha\beta}$ replaced by $g^{R}_{\alpha\beta}$.

(2) Equation (6.30) is obtained by combining Eq. (6.28b), which can be rewritten as

$$\frac{\delta \bar{g}^{\xi\eta}}{\delta g^{R}_{\alpha\beta}}\frac{\delta}{\delta \bar{g}^{\xi\eta}}\Gamma[\bar{g}^{\lambda\sigma},g^{R}_{\alpha\beta}]+\frac{\delta}{\delta g^{R}_{\alpha\beta}}\Gamma[\bar{g}^{\lambda\sigma},g^{R}_{\alpha\beta}]=0\ ,$$

with Eq. (6.29), which implies the vanishing of the first term on the left-hand side of the above equation.

(3) In Eq. (6.58), $\langle h_{\theta\tau}(0)\rangle$ is a shorthand for $\langle h_{\theta\tau}(0)\rangle_J$, with $J^{\xi\eta}[\bar{g}_{\alpha\beta}]$ the external source current.

(4) In deriving Eq. (6.59), use has been made of the identity

$$0=\int d[\,]e^{i\bar{S}}V^{\mu\nu}(x)V_2^{\alpha\beta}(y)\ ,$$

which follows from Eq. (6.58b) and the fact that $V_2^{\alpha\beta}$ is linear in $h_{\lambda\sigma}$. From this identity we then get

$$\begin{aligned}0&=\langle\mathscr{T}([V_1(x)+V_2(x)]V_2(0))\rangle_0\\&\Longrightarrow\langle\mathscr{T}([V_1(x)+V_2(x)][V_1(0)+V_2(0)])\rangle_0\\&=\langle\mathscr{T}(V_1(x)V_1(0))\rangle_0-\langle\mathscr{T}(V_2(x)V_2(0))\rangle_0\ .\end{aligned}$$

I wish to thank A. Zee for comments on Sec. VI.

In the references, the paper of Brout, Englert, and Gunzig (1978) appeared in Ann. Phys. (N.Y.), not in Ann. Phys. (Paris). The citation of Utiyama and DeWitt (1962) in Sec. VI.D should also refer to DeWitt (1950) [DeWitt, B.S., 1950, Ph.D. thesis (Harvard University), unpublished].

Excerpt from S. L. Adler, Theory of Static Quark Forces, *Phys. Rev. D.* **18**, 411–434 (1978).

APPENDIX A: SCALAR PROPAGATOR CONSTRUCTION

I construct in this appendix the scalar propagator $\Delta^{ab}(x,y)$ defined in Eq. (81) of the text, which satisfies

$$
\begin{aligned}
D_x^\mu D_x^\mu \Delta^{ab}(x,y) &= (D_x^0 D_x^0 + D_x^j D_x^j)\Delta^{ab}(x,y) \\
&= -\delta^{ab}\delta^3(x-y), \\
D_x^0 \vec{w}(x) &= \vec{\lambda}_0(x) \times \vec{w}(x), \\
D_x^j \vec{w}(x) &= \left[\frac{\partial}{\partial x^j} + \vec{b}_0^j(x)\times\right]\vec{w}(x),
\end{aligned} \tag{A1}
$$

$$\lambda_0^a(x) = \frac{x^a}{r^2}(1 - r\coth r),$$

$$b_0^{al}(x) = \frac{\epsilon^{alj}x^j}{r^2}\left(1 - \frac{r}{\sinh r}\right),$$

where I have set $\kappa = 1$. To change to general κ one simply uses the scaling law

$$\Delta^{ab}(x, y)_{\text{general}\,\kappa} \equiv \Delta^{ab}(x, y; \kappa)$$
$$= \kappa\Delta^{ab}(\kappa x, \kappa y; 1). \tag{A2}$$

The reversal in sign of $\vec{\lambda}_0$ as compared with Eq. (49) [which I have made because the sign in Eq. (A1) corresponds to the convention I used in my calculations] has no effect on Δ^{ab}, since $D_x^\mu D_x^\mu$ is even in $\vec{\lambda}_0$. As in the vector propagator calculation in the text, I make extensive use of the results of Brown *et al.*[20] for propagators in pseudoparticle fields. The first step of the calculation, following Manton,[31] is to make a complex gauge transformation which changes the potentials from λ_0^a, b_0^{al} of Eq. (A1) to $\bar{\lambda}_0^a, \bar{b}_0^{al}$, with

$$\bar{\lambda}_0^a = \lambda_0^a = \frac{x^a}{r^2}(1 - r\coth r),$$
$$\bar{b}_0^{al} = \frac{\epsilon^{alj}x^j}{r^2}(1 - r\coth r) + i\delta^{la}. \tag{A3}$$

Introducing a matrix $M^{a\bar{a}}(x)$ given by

$$M^{a\bar{a}}(x) = \cosh r\left(\delta^{a\bar{a}} - \frac{x^a x^{\bar{a}}}{r^2}\right)$$
$$-i\sinh r\,\epsilon^{a\bar{a}l}\frac{x^l}{r} + \frac{x^a x^{\bar{a}}}{r^2}, \tag{A4}$$
$$M^{a\bar{a}}(x)M^{b\bar{a}}(x) = \delta^{ab},$$

it is straightforward to verify that

$$\frac{\partial}{\partial x^j}M^{a\bar{a}}(x) = -\epsilon^{abc}b_0^{bj}M^{c\bar{a}} + M^{a\bar{e}}\epsilon^{\bar{e}\bar{b}\bar{a}}\bar{b}_0^{\bar{b}j}, \tag{A5}$$

which implies that

$$D_x^j M^{a\bar{a}}(x) = M^{a\bar{a}}(x)\bar{D}_x^j. \tag{A6}$$

So once we have obtained the scalar propagator $\bar{\Delta}^{\bar{a}\bar{b}}(x)$ satisfying

$$\bar{D}_x^\mu \bar{D}_x^\mu \bar{\Delta}^{\bar{a}\bar{b}}(x, y) = -\delta^{\bar{a}\bar{b}}\delta^3(x - y), \tag{A7}$$

we get the desired propagator Δ^{ab} by transforming both SU(2) indices with the matrix M,

$$\Delta^{ab}(x, y) = M^{a\bar{a}}(x)M^{b\bar{b}}(y)\bar{\Delta}^{\bar{a}\bar{b}}(x, y). \tag{A8}$$

From this point on I will work exclusively with the gauge-transformed potentials of Eq. (A3). For notational convenience I will drop all bars, but it should be kept in mind that I am now constructing the propagator $\bar{\Delta}$ in the new gauge, not the final propagator Δ given by Eq. (A8). The advantage of the potentials of Eq. (A3) is that they take the form used by Brown *et al.* as the starting point for their analysis,

$$A^{a\mu}(x) = (\lambda_0^a, b_0^{al}) = -\eta^{(-)\mu\nu a}\partial^\nu \ln\pi(x),$$
$$\pi(x) = e^{ix^0}\frac{\sinh r}{r}, \quad \partial^0 = \frac{\partial}{\partial x^0}, \quad \partial^j = \frac{\partial}{\partial x^j}, \tag{A9}$$
$$\eta^{(-)\mu\nu a} = -\eta^{(-)\nu\mu a}, \quad \eta^{(-)kla} = \epsilon^{kla}, \quad \eta^{(-)k0a} = -\delta^{ka}.$$

Note that although $\pi(x)$ depends on x^0, the potential $A^{a\mu}(x)$ depends only on the spatial components x^j of x. Hence if I define a Euclidean time-dependent propagator $\Delta^{ab}(\vec{x}, \vec{y}, x^0, y^0)$ by

$$D_x^\mu D_x^\mu \Delta^{ab}(\vec{x}, \vec{y}, x^0, y^0) = -\delta^{ab}\delta^4(x - y),$$
$$D_x^0\vec{w}(x) = \left[\frac{\partial}{\partial x^0} + \vec{\lambda}_0(x)\times\right]\vec{w}(x), \tag{A10}$$

then it will actually depend only on the time difference $\lambda = x^0 - y^0$, and the desired propagator $\Delta^{ab}(\vec{x}, \vec{y})$ is obtained by integrating over the time difference,

$$\Delta^{ab}(\vec{x}, \vec{y}) = \int_{-\infty}^{\infty} d\lambda\, \Delta^{ab}(\vec{x}, \vec{y}, \lambda). \tag{A11}$$

The final observation needed, in order to make contact with the work of Brown *et al.*, is that $\pi(x)$ can be written as a contour integral,

$$\pi(x) = \frac{e^{ix_0}\sinh r}{r}$$
$$= -\frac{1}{2\pi}\int_{-\infty - iK}^{\infty - iK} ds\, e^{is}\frac{1}{r^2 + (x^0 - s)^2}, \quad K > r. \tag{A12}$$

I now will list a number of results from the analysis of Brown *et al.*, with occasional small changes in notation. Brown *et al.* construct the general scalar, isovector propagator $\Delta^{ab}(x, y, x^0, y^0)$ satisfying Eq. (A10) for potentials $A^{a\mu}(x) = -\eta^{(-)\mu\nu a}\partial^\nu \ln\pi(x)$ representing a general N-pseudoparticle (instanton) configuration

$$\pi(x) = (1) + \sum_s \frac{\rho_s^2}{x_s^2}, \quad x_s \equiv x - z_s. \tag{A13}$$

Their result takes the form of a sum of two pieces (with x, y Euclidean four-vectors)

$$\Delta^{ab}(x,y) = \Delta^{ab}(x,y)^{(1)} + \Delta^{ab}(x,y)^{(2)}. \tag{A14}$$

The first piece is constructed in terms of spin-$\frac{1}{2}$ propagators by the recipe

$$\begin{aligned}
&\Delta^{ab}(x,y)^{(1)} = \frac{U^{ab}(x,y)}{4\pi^2(x-y)^2\pi(x)\pi(y)},\\
&U^{ab}(x,y) = \tfrac{1}{2}\mathrm{tr}[\tau^a F^{(+)}(x,y)\tau^b F^{(+)}(y,x)],\\
&F^{(+)}(x,y) = (1) + \sum_s \rho_s^2 \frac{\tau\cdot x_s}{x_s^2}\frac{\tau^\dagger\cdot y_s}{y_s^2},
\end{aligned} \tag{A15}$$

$\tau^\mu = (i, \tau^j)$, $\tau^{\dagger\mu} = (-i, \tau^j)$,

$\tau^j = \mathrm{SU}(2)$ Pauli matrices, $x_s = x - z_s$, $y_s = y - z_s$.

The second piece has the form

$$\begin{aligned}
&\Delta^{ab}(x,y)^{(2)} = \frac{C_{ab}(x,y)}{4\pi^2\pi(x)\pi(y)},\\
&C_{ab}(x,y) = \sum_{r,s,u,v} \Phi^{(+)}_{rsa}(x) C_{rs,uv} \Phi^{(+)}_{uvb}(y),\\
&\Phi^{(+)}_{rsa}(x) = \frac{\rho_r\rho_s}{x_r^2 x_s^2}\eta^{(-)\mu\nu a} x_r^\mu x_s^\nu,\\
&C_{rs,uv} = \frac{\delta_{ru}\delta_{sv} - \delta_{rv}\delta_{su}}{(z_r-z_s)^2} - \frac{1}{(z_r-z_s)^2}\\
&\qquad\times\left(\frac{\rho_r\rho_u}{\rho_s\rho_v}h_{sv} - \frac{\rho_r\rho_v}{\rho_s\rho_u}h_{su} - \frac{\rho_s\rho_u}{\rho_r\rho_v}h_{rv} + \frac{\rho_s\rho_v}{\rho_r\rho_u}h_{ru}\right)\frac{1}{(z_u-z_v)^2},
\end{aligned} \tag{A16}$$

and with the numbers[32] h_{sv} determined by the matrix inversion problem

$$\sum_t \frac{g_{st}h_{tv}}{\rho_t\rho_v} = \delta_{sv}, \tag{A17a}$$

$$g_{st} = \left[(1) + \sum_{r\neq s}\frac{\rho_r^2}{(z_r-z_s)^2}\right]\delta_{st} - \frac{\rho_s\rho_t}{(z_s-z_t)^2}(1-\delta_{st}). \tag{A17b}$$

To make use of these rather formidable looking equations, I note that Eq. (A13) becomes identical to Eq. (A12) under the substitutions

$$\begin{aligned}
&(1)\to 0,\\
&z_s\to(s,0),\quad x_s^2\to(x^0-s)^2+\vec{x}^2,\\
&\sum_s \to \int ds,\quad \rho_s\to -(1/2\pi)e^{is},
\end{aligned} \tag{A18}$$

so that the Sommerfield-Prasad solution is in effect a continuum of complex instantons. Corresponding to the substitution $(1)\to 0$, the terms (1) in Eqs. (A15) and (A17) must also be deleted. The transition from sums to integrals can be made with no ambiguity in $\Delta^{ab}(x,y)^{(1)}$, giving (recall that $\lambda = x^0 - y^0$)

$$\begin{aligned}
\Delta^{ab}(\vec{x},\vec{y})^{(1)} &\equiv \int_{-\infty}^{\infty} d\lambda\, \Delta^{ab}(\vec{x},\vec{y},\lambda)^{(1)}\\
&= \int_{-\infty}^{\infty} d\lambda \frac{1}{4\pi^2[(\vec{x}-\vec{y})^2+\lambda^2]}\frac{|\vec{x}|}{\sinh|\vec{x}|}\frac{|\vec{y}|}{\sinh|\vec{y}|} e^{-ix^0-iy^0}\tfrac{1}{2}\mathrm{tr}[\tau^a F^{(+)}(x,y)\tau^b F^{(+)}(y,x)],\\
F^{(+)}(x,y) &= -\frac{1}{2\pi}\int ds\, e^{is}\frac{\vec{\tau}\cdot\vec{x}+i(x^0-s)}{\vec{x}^2+(x^0-s)^2}\frac{\vec{\tau}\cdot\vec{y}-i(y^0-s)}{\vec{y}^2+(y^0-s)^2}.
\end{aligned} \tag{A19}$$

Making a shift $s\to s+x^0$ in the integration over s in $F^{(+)}(x,y)$ and a shift $t\to t+x^0$ in the corresponding integration over t in $F^{(+)}(y,x)$, gives

$$\begin{aligned}
\Delta^{ab}(x,y)^{(1)} = \frac{1}{(2\pi)^4}\frac{|\vec{x}|}{\sinh|\vec{x}|}\frac{|\vec{y}|}{\sinh|\vec{y}|}\int ds\,e^{is}\int dt\,e^{it}\int_{-\infty}^{\infty}\frac{d\lambda\,e^{i\lambda}}{(\vec{x}-\vec{y})^2+\lambda^2}\\
\times\tfrac{1}{2}\mathrm{tr}\left[\frac{\vec{\tau}\cdot\vec{x}+it}{\vec{x}^2+t^2}\tau^a\frac{\vec{\tau}\cdot\vec{x}-is}{\vec{x}^2+s^2}\frac{\vec{\tau}\cdot\vec{y}+i(s+\lambda)}{\vec{y}^2+(s+\lambda)^2}\tau^b\frac{\vec{\tau}\cdot\vec{y}-i(t+\lambda)}{\vec{y}^2+(t+\lambda)^2}\right].
\end{aligned} \tag{A20}$$

Now make, in the order indicated, the following changes of variables:

$$\begin{aligned}
&\text{(i)}\quad \lambda\to z-\tfrac{1}{2}(s+t) = z-w,\\
&\text{(ii)}\quad w=\tfrac{1}{2}(s+t),\quad v=\tfrac{1}{2}(s-t),\quad ds\,dt = 2\,dw\,dv.
\end{aligned} \tag{A21}$$

This gives as the final result the following symmetrical-looking formula:

$$\Delta^{ab}(x,y)^{(1)}=\frac{2}{(2\pi)^4}\frac{|\vec{x}|}{\sinh|\vec{x}|}\frac{|\vec{y}|}{\sinh|\vec{y}|}\int_{-\infty}^{\infty}dv\int_{-\infty-iK}^{\infty-iK}dw\int_{-\infty-iK}^{\infty-iK}dz\,\frac{e^{iw}e^{iz}}{(\vec{x}-\vec{y})^2+(z-w)^2}$$
$$\times\tfrac{1}{2}\mathrm{tr}\left[\frac{\vec{\tau}\cdot\vec{x}+i(w-v)}{\vec{x}^2+(w-v)^2}\tau^a\frac{\vec{\tau}\cdot\vec{x}-i(w+v)}{\vec{x}^2+(w+v)^2}\frac{\vec{\tau}\cdot\vec{y}+i(z+v)}{\vec{y}^2+(z+v)^2}\tau^b\frac{\vec{\tau}\cdot\vec{y}-i(z-v)}{\vec{y}^2+(z-v)^2}\right].\tag{A22}$$

Turning next to the second piece, I note that time-translation invariance implies that $h_{sv}=h(s-v)$. Anticipating the fact that only $H=\sum_v h(s-v)$ is needed, I proceed first to extract this quantity from the matrix inversion problem of Brown *et al.* stated in Eq. (A17). Because the expressions of Eqs. (A16) and (A17) contain singular factors $(z_u-z_v)^{-2}$, etc., it is necessary to separate the various integration contours r,s,u,v by small imaginary displacements. In order to do this in a way which preserves the validity of various algebraic operations used by Brown *et al.* in getting their solution,[33] it is necessary to symmetrize over all possible "stacking orders" of the contours on the complex plane, a procedure which will eventually lead to the appearance of principal-value integrals in the answer. Summing over v in Eq. (A17a) gives

$$\left[\sum_v h(t-v)\right]\left[\sum_t\frac{g_{st}}{\rho_t}\right]=\sum_v\rho_v\delta_{sv}=\rho_s\,.\tag{A23}$$

Dropping the (1) in the expression for g_{st} in Eq. (A17b) gives

$$\sum_t\frac{g_{st}}{\rho_t}=\rho_s\sum_{r\neq s}\frac{\rho_r^2/\rho_s^2-1}{(z_r-z_s)^2}\,,\tag{A24}$$

so that

$$H=\sum_v h(t-v)=\left[\sum_{r\neq s}\frac{\rho_r^2/\rho_s^2-1}{(z_r-z_s)^2}\right]^{-1}.\tag{A25}$$

As a consistency check, note that if we multiply Eq. (A23) by ρ_s and sum, we get

$$H\sum_{s,t}\frac{g_{st}}{\rho_t}\rho_s=H\sum_{r\neq s}\frac{\rho_r^2-\rho_s^2}{(z_r-z_s)^2}$$
$$=0=\sum_s\rho_s^2\,,\tag{A26}$$

but in the continuum limit

$$\sum_s\rho_s^2\to-\frac{1}{2\pi}\int_{-\infty-iK}^{\infty-iK}ds\,e^{is}=0\,,\tag{A27}$$

so that Eq. (A26) is in fact satisfied. Passing to the limit in Eq. (A25), and remembering that we must average over the cases where the r contour goes over and under s, we get

$$H=\left[\mathrm{P}\int_{-\infty-iK}^{\infty-iK}dr\,\frac{e^{i(r-s)}-1}{(r-s)^2}\right]^{-1}=-\frac{1}{\pi}\,.\tag{A28}$$

The final step in the calculation is to make the transition froms sums to integrals in Eq. (A16), bearing in mind the necessity of symmetrizing over the ordering of integration contours. Noting that

$$\eta^{(-)\mu\nu a}x_{r\mu}x_{s\nu}=-x^a(r-s)\,,\tag{A29}$$

we get from Eq. (A16) (again with $\lambda=x^0-y^0$)

$$\Delta^{ab}(\vec{x},\vec{y})^{(2)}\equiv\int_{-\infty}^{\infty}d\lambda\,\Delta^{ab}(\vec{x},\vec{y},\lambda)^{(2)}$$
$$=\frac{1}{(2\pi)^3}\frac{|\vec{x}|}{\sinh|\vec{x}|}\frac{|\vec{y}|}{\sinh|\vec{y}|}x^a y^b$$
$$\times\int_{-\infty}^{\infty}d\lambda\,e^{-i(x^0+y^0)}\left\{\frac{2}{(2\pi)^2}\int dr\,e^{ir}\int ds\,e^{is}\frac{1}{[\vec{x}^2+(x^0-r)^2][\vec{x}^2+(x^0-s)^2][\vec{y}^2+(y^0-r)^2][\vec{y}^2+(y^0-s)^2]}\right.$$
$$-\frac{4}{(2\pi)^2}\int dr\,e^{ir}\int du\,e^{iu}\int\frac{ds\,dv\,h(s-v)}{(r-s)(u-v)}$$
$$\left.\times\frac{1}{[\vec{x}^2+(x^0-r)^2][\vec{x}^2+(x^0-s)^2][\vec{y}^2+(y^0-u)^2][\vec{y}^2+(y^0-v)^2]}\right\}.\tag{A30}$$

Again it is necessary to make, in the order indicated, the following changes of variables;

First term in { }:

$$\begin{aligned}&\text{(i)}\ r\to r+x^0,\quad s\to s+x^0\ ,\\ &\text{(ii)}\ \lambda\to z-\tfrac{1}{2}(r+s)=z-w\ ,\\ &\text{(iii)}\ w=\tfrac{1}{2}(r+s),\quad v=\tfrac{1}{2}(r-s),\quad drds=2dwdv\ .\end{aligned}\tag{A31a}$$

Second term in { }:

$$\begin{aligned}&\text{(i)}\ r\to r+x^0,\quad s\to s+x^0,\quad u\to u+x^0,\quad v\to v+x^0\ ,\\ &\text{(ii)}\ \lambda\to z_1-\tfrac{1}{2}(u+v)\ ,\\ &\text{(iii)}\ z_2=\tfrac{1}{2}(r+s),\quad w=\tfrac{1}{2}(u+v),\quad v_2=\tfrac{1}{2}(r-s),\quad v_1=\tfrac{1}{2}(u-v),\\ &\qquad drds=2dz_2dv_2,\quad dudv=2dwdv_1\ .\end{aligned}\tag{A31b}$$

After these transformations, the only place where w appears is in

$$\int dw\, h(z_2-v_2+v_1-w)=-1/\pi\ ,\tag{A32}$$

giving as the final answer

$$\begin{aligned}\Delta^{ab}(\vec{x},\vec{y})^{(2)}=&\frac{4}{(2\pi)^4}\frac{|\vec{x}|}{\sinh|\vec{x}|}\frac{|\vec{y}|}{\sinh|\vec{y}|}x^ay^b\\ &\times\Bigg\{\int_{-\infty}^{\infty}dv\int_{-\infty-iK}^{\infty-iK}\frac{dw\,e^{iw}}{[\vec{x}^2+(w-v)^2][\vec{x}^2+(w+v)^2]}\int_{-\infty-iK}^{\infty-iK}\frac{dz\,e^{iz}}{[\vec{y}^2+(z-v)^2][\vec{y}^2+(z+v)^2]}\\ &+\frac{1}{\pi}\,\mathrm{P}\int_{-\infty}^{\infty}dv_2\frac{e^{iv_2}}{v_2}\int_{-\infty-iK}^{\infty-iK}\frac{dz_2\,e^{iz_2}}{[\vec{x}^2+(z_2-v_2)^2][\vec{x}^2+(z_2+v_2)^2]}\\ &\qquad\times\mathrm{P}\int_{-\infty}^{\infty}dv_1\frac{e^{iv_1}}{v_1}\int_{-\infty-iK}^{\infty-iK}\frac{dz_1\,e^{iz_1}}{[\vec{y}^2+(z_1-v_1)^2][\vec{y}^2+(z_1+v_1)^2]}\Bigg\}.\end{aligned}\tag{A33}$$

Although it took a more involved argument to extract Eq. (A33) from the work of Brown *et al.* than was needed to get Eq. (A22), the evaluation of the contour integrals appearing in Eq. (A33) is relatively easy. Writing $x=|\vec{x}|$, $y=|\vec{y}|$, the answer is

$$\begin{aligned}\Delta^{ab}(\vec{x},\vec{y})^{(2)}=&\frac{1}{4\pi}\frac{x^a}{\sinh x}\frac{y^b}{\sinh y}\\ &\times\Bigg\{\frac{1}{xy}\left[\cosh x\cosh y-\frac{1}{2}\left(\frac{\sinh x}{x}\cosh y+\frac{\sinh y}{y}\cosh x\right)\right]\\ &+\frac{1}{4}\left(\frac{1}{x^2}+\frac{1}{y^2}\right)\left(\frac{\sinh(x-y)}{x-y}-\frac{\sinh(x+y)}{x+y}\right)-\frac{1}{xy}\left(\cosh x-\frac{\sinh x}{x}\right)\left(\cosh y-\frac{\sinh y}{y}\right)\Bigg\}\\ =&\frac{1}{4\pi}\frac{x^a}{\sinh x}\frac{y^b}{\sinh y}\Bigg\{\frac{1}{2xy}\left(\frac{\sinh x}{x}\cosh y+\frac{\sinh y}{y}\cosh x\right)-\frac{\sinh x}{x^2}\frac{\sinh y}{y^2}\\ &+\frac{1}{4}\left(\frac{1}{x^2}+\frac{1}{y^2}\right)\left[\frac{\sinh(x-y)}{x-y}-\frac{\sinh(x+y)}{x+y}\right]\Bigg\}\ .\end{aligned}\tag{A34b}$$

The fact that the final term in Eq. (A34a), which comes from the product of principal-value integrals in Eq. (A33), cancels away the leading large-y asymptotic behavior of the first three terms is a check that the limiting argument leading to Eq. (A28) has been carried out correctly. The evaluation of $\Delta^{ab}(\vec{x},\vec{y})^{(1)}$, in which $\vec{x}$ and $\vec{y}$ dependences are highly correlated, involves straightforward but very lengthy computations, on which I am now working.

Excerpt from S. L. Adler, Classical Quark Statics, *Phys. Rev. D.* 19, 1168–1187 (1979).

APPENDIX A: PROPAGATOR FORMULAS

In giving formulas for the scalar propagator Δ^{ab} in the Prasad-Sommerfield background field, it is convenient to set $\kappa = 1$; to change to general κ one uses the scaling law

$$\begin{aligned}\Delta^{ab}(x,y)_{\text{general }\kappa} &\equiv \Delta^{ab}(x,y;\kappa)\\ &= \kappa\Delta^{ab}(\kappa x,\kappa y;1)\,.\end{aligned} \tag{A1}$$

Writing

$$\Delta^{ab}(x,y;1) = \frac{1}{4\pi}\,\frac{x}{\sinh x}\,\frac{y}{\sinh y}\,\Sigma^{ab}, \tag{A2}$$

I find the following expression for Σ^{ab}:

$$\Sigma^{ab} = \sum_{i=1}^{5} \sigma_i^{ab}(x,y)\lambda_i(x,y)\,,$$

$$\begin{aligned}
\sigma_1^{ab} &= \delta^{ab} + \frac{\vec{x}\cdot\vec{y}}{x^2y^2}\,x^a y^b - \frac{x^a x^b}{x^2} - \frac{y^a y^b}{y^2}\,,\\
\sigma_2^{ab} &= x^a y^b\,,\\
\sigma_3^{ab} &= x^b y^a - \delta^{ab}\vec{x}\cdot\vec{y}\,,\\
\sigma_4^{ab} &= \frac{x^a}{x^2}\left(x^b - y^b\,\frac{\vec{x}\cdot\vec{y}}{y^2}\right),\\
\sigma_5^{ab} &= \frac{y^b}{y^2}\left(y^a - x^a\,\frac{\vec{x}\cdot\vec{y}}{x^2}\right),
\end{aligned} \tag{A3}$$

$$\begin{aligned}
\lambda_1 &= \frac{1}{2\Delta}\left[f_2(z_{++}) + f_2(z_{--}) + f_2(z_{+-}) + f_2(z_{-+})\right]\\
&= \frac{2}{\Delta}\int_0^1 d\alpha(1-\alpha)e^{-\alpha\Delta}\cosh\alpha x\cosh\alpha y\,,\\
\lambda_2 &= \frac{1}{x^2y^2}\left(\frac{\cosh x\cosh y - e^{-\Delta}}{\Delta} - \frac{\sinh x}{x}\,\frac{\sinh y}{y}\right),\\
\lambda_3 &= \frac{1}{2xy\Delta}\left[-f_2(z_{++}) - f_2(z_{--}) + f_2(z_{+-}) + f_2(z_{-+})\right]\\
&= -\frac{2}{\Delta}\int_0^1 d\alpha(1-\alpha)e^{-\alpha\Delta}\,\frac{\sinh\alpha x}{x}\,\frac{\sinh\alpha y}{y}\,,\\
\lambda_4 &= \frac{1}{2x\Delta}\{e^{x}[f_1(z_{-+}) + f_1(z_{--})]\\
&\qquad - e^{-x}[f_1(z_{+-}) + f_1(z_{++})]\}\\
&= \frac{2}{\Delta}\int_0^1 d\alpha\, e^{-\alpha\Delta}\cosh\alpha y\,\frac{\sinh(1-\alpha)x}{x}\,,\\
\lambda_5 &= \frac{1}{2y\Delta}\{e^{y}[f_1(z_{+-}) + f_1(z_{--})]\\
&\qquad - e^{-y}[f_1(z_{-+}) + f_1(z_{++})]\}\\
&= \frac{2}{\Delta}\int_0^1 d\alpha\, e^{-\alpha\Delta}\cosh\alpha x\,\frac{\sinh(1-\alpha)y}{y}\,,\\
\Delta &= |\vec{x} - \vec{y}|\,,\\
z_{++} &= x + y - \Delta\,,\quad z_{+-} = x - y - \Delta\,,\\
z_{-+} &= -x + y - \Delta\,,\quad z_{--} = -x - y - \Delta\,,\\
f_1(z) &= \frac{e^z - 1}{z}\,,\quad f_2(z) = \frac{e^z - 1 - z}{z^2}\,.
\end{aligned} \tag{A4}$$

It is easy to verify that, despite the factors x^{-2} and y^{-2} in the σ's, Eq. (A3) is in fact analytic near $x = 0$ and near $y = 0$. Near $\vec{x} = \vec{y}$, Eq. (A2) has the expected short-distance singularity

$$\Delta^{ab}(x,y;1) \underset{\Delta\to 0}{\sim} \frac{1}{4\pi}\,\frac{\delta^{ab}}{\Delta} + O(1)\,, \tag{A5}$$

while the limiting behavior for large y, with x fixed, is

$$\Delta^{ab}(x,y;1) \underset{y\to\infty}{\sim} \frac{1}{4\pi}\,\frac{x^a}{x}\left(\coth x - \frac{1}{x}\right)\frac{y^b}{y^2} + O\!\left(\frac{1}{y^2}\right), \tag{A6}$$

which has $b_0^{a0}(x)$ as the x-dependent factor. The expression for the differential operator $D_x^{\mu}D_x^{\mu}$ used in verifying the propagator differential equation is

$$\begin{aligned}
&\left(\frac{\sinh x}{x}\,D_{0x}^{\sigma}D_{0x}^{\sigma}\,\frac{x}{\sinh x}\,\vec{\phi}\right)^a = \left(\frac{\partial}{\partial x^j}\right)^2\phi^a + C_1\phi^a + C_2\hat{x}^j\frac{\partial}{\partial x^j}\phi^a + C_3\hat{x}^a\hat{x}^j\phi^j + C_4\hat{x}^j\frac{\partial}{\partial x^a}\phi^j + C_5\hat{x}^a\frac{\partial}{\partial x^j}\phi^j\,,\\
&C_1 = \frac{2}{x}\left(\frac{1}{\sinh x} - \coth x\right),\quad C_2 = 2\left(\frac{1}{x} - \coth x\right),\quad C_3 = 1 + C_1\,,\\
&C_4 = -2\left(\frac{1}{x} - \frac{1}{\sinh x}\right),\quad C_5 = -C_4\,.
\end{aligned} \tag{A7}$$

The projected covariant derivative of the scalar Green's function, defined by Eq. (46) of the text, is given

by the following formulas:

$$\Delta^j(x,y;1)^{Lb}=\frac{1}{4\pi}\frac{y}{\sinh y}\sum_{j=1}^{5}\sigma^{jb}(x,y)\tau_j(x,y),$$

$$\begin{aligned}
\tau_1(x,y)&=\frac{1}{\sinh x}\left(\lambda_4-\frac{\vec{x}\cdot\vec{y}}{\Delta}\frac{\partial\lambda_4}{\partial\Delta}\right)-\frac{x}{\sinh^2 x}\lambda_1,\\
\tau_2(x,y)&=\left(\frac{2}{\sinh x}-\frac{x\cosh x}{\sinh^2 x}\right)\lambda_2+\frac{x}{\sinh x}\frac{\partial\lambda_2}{\partial x}+\frac{x^2-\vec{x}\cdot\vec{y}}{\Delta\sinh x}\frac{\partial\lambda_2}{\partial\Delta},\\
\tau_3(x,y)&=-\frac{1}{\Delta\sinh x}\frac{\partial\lambda_4}{\partial\Delta}-\frac{x}{\sinh^2 x}\lambda_3,\\
\tau_4(x,y)&=\left(\frac{1}{\sinh x}-\frac{x\cosh x}{\sinh^2 x}\right)\lambda_4+\frac{x}{\sinh x}\frac{\partial\lambda_4}{\partial x}+\frac{x^2-\vec{x}\cdot\vec{y}}{\Delta\sinh x}\frac{\partial\lambda_4}{\partial\Delta},\\
\tau_5(x,y)&=-\frac{x^2y^2}{\Delta\sinh x}\frac{\partial\lambda_2}{\partial\Delta}-\frac{x}{\sinh^2 x}\lambda_5,
\end{aligned}\tag{A8}$$

with

$$\begin{aligned}
\frac{\partial\lambda_4}{\partial\Delta}&=-\left(1+\frac{1}{\Delta}\right)\lambda_4+\frac{1}{2x\Delta}\{e^x[f_2(z_{-+})+f_2(z_{--})]-e^{-x}[f_2(z_{+-})+f_2(z_{++})]\},\\
\frac{\partial\lambda_4}{\partial x}&=-\frac{1}{x}\lambda_4+\frac{1}{2x\Delta}\{e^x[f_2(z_{-+})+f_2(z_{--})]+e^{-x}[f_2(z_{+-})+f_2(z_{++})]\},\\
\frac{\partial\lambda_2}{\partial\Delta}&=\frac{1}{x^2y^2}\left(-\frac{\cosh x\cosh y-e^{-\Delta}}{\Delta^2}+\frac{e^{-\Delta}}{\Delta}\right),\\
\frac{\partial\lambda_2}{\partial x}&=-\frac{2}{x}\lambda_2+\frac{1}{x^2y^2}\left[\frac{\sinh x\cosh y}{\Delta}-\left(\frac{\cosh x}{x}-\frac{\sinh x}{x^2}\right)\frac{\sinh y}{y}\right].
\end{aligned}\tag{A9}$$

In the region $y\gg x$, $y\gg 1$ the following formula is useful:

$$\begin{aligned}
\Delta^j(x,y;1)^{Lb}&=\frac{1}{4\pi}\left[x^jy^b\tau_2^s+\frac{y^b}{y^2}\left(y^j-x^j\frac{\vec{x}\cdot\vec{y}}{x^2}\right)\tau_5^s\right]+O(e^{-y}),\\
\tau_2^s&=\frac{1}{y^2}\frac{1}{x^3}-\frac{1}{\Delta y}\frac{1}{x\sinh^2 x}-\frac{x^2-\vec{x}\cdot\vec{y}}{\Delta^3 y}\frac{\cosh x}{x^2\sinh x},\\
\tau_5^s&=\frac{y}{\Delta^3}\frac{\cosh x}{\sinh x}-\frac{x}{\sinh^2 x}\frac{1}{\Delta}\left(\frac{1}{y+\Delta-x}+\frac{1}{y+\Delta+x}\right).
\end{aligned}\tag{A10}$$

Relaxation methods for gauge field equilibrium equations

Stephen L. Adler

The Institute for Advanced Study, Princeton, New Jersey 08540

Tsvi Piran

The Institute for Advanced Study, Princeton, New Jersey 08540
and Racah Institute of Physics, The Hebrew University, Jerusalem, Israel

This article gives a pedagogical introduction to relaxation methods for the numerical solution of elliptic partial differential equations, with particular emphasis on treating nonlinear problems with δ-function source terms and axial symmetry, which arise in the context of effective Lagrangian approximations to the dynamics of quantized gauge fields. The authors present a detailed theoretical analysis of three models which are used as numerical examples: the classical Abelian Higgs model (illustrating charge screening), the semiclassical leading logarithm model (illustrating flux confinement within a free boundary or "bag"), and the axially symmetric Bogomol'nyi-Prasad-Sommerfield monopoles (illustrating the occurrence of topological quantum numbers in non-Abelian gauge fields). They then proceed to a self-contained introduction to the theory of relaxation methods and allied iterative numerical methods and to the practical aspects of their implementation, with attention to general issues which arise in the three examples. The authors conclude with a brief discussion of details of the numerical solution of the models, presenting sample numerical results.

CONTENTS

I. INTRODUCTION

Over the last few years, numerical methods have played an increasingly important role in theoretical physics. Their prominence is attributable both to improvements in computers and decreased computational costs, and to the increased attention of theorists to nonlinear, nonperturbative problems in quantum field theory, for which purely analytical methods are inadequate. In treating quantum field theories computationally, two strategies are possible. The first, which has been intensively pursued recently, is to set up a discrete lattice analog of the full quantum field theory, and then to numerically evaluate the Feynman path integral by Monte Carlo techniques. This method has the advantage of giving results which can in principle be made as accurate as desired. Nonetheless, the necessity of using a four-dimensional computational lattice and of generating a large ensemble of field configurations makes simulation very costly, and in practice this has been a severe limiting factor. A second strategy is to first make analytic approximations, which replace the field theoretic problem by a classical variational problem involving an effective Lagrangian functional, leading to a system of partial differential equations which are then solved numerically. This approach is necessarily approximate, since exact knowledge of the effective Lagrangian is not possible without an exact evaluation of the Feynman path integral. However, the second strategy has the advantages that symmetries of the physical problem can be exploited to reduce the dimensionality of the computational problem, and that only a single equilibrium field configuration need be generated, permitting the study of very large computational lattices even on small computers. We believe that the two strategies are, in a sense, complementary; eventually, simulations may be used to infer the form of effective Lagrangians, which can then be used to analyze large classes of problems of physical interest.

Our aim in this article is to give a pedagogical review of the numerical analysis methods required by the second strategy. Assuming that an approximate nonlinear classical effective Lagrangian has been given, we show how relaxation methods can be used to solve the partial differential equations which govern the equilibrium field configurations. We focus on problems which arise in gauge field theories of current interest and in Sec. II introduce and

analytically characterize three nonlinear models which will be studied as illustrative examples. In Sec. III we give a self-contained introduction to the theory of relaxation methods and to practical aspects of their implementation, with special emphasis on treating nonlinear problems with singular (δ-function) source terms. Readers primarily interested in numerical analysis can proceed directly to Sec. III, after reading only the brief survey and theoretical discussion of Secs. II.A and II.B. In Sec. IV we give specific details of the application of the methods of Sec. III to the models of Secs. II.C, II.D, and II.E, together with a small sampling of numerical results. Certain technical details of the analytical structure of the three models are described in the Appendixes.

II. THEORETICAL ANALYSIS OF MODELS TO BE STUDIED

A. Introduction

In this section, we give a self-contained theoretical analysis of the models which later on will be studied numerically. All of the models discussed below describe time-independent three-dimensional problems with a rotational symmetry axis, and hence lead to two-dimensional computational problems in cylindrical coordinates. Our primary focus will be on the statics of classical charges in nonlinear Abelian and non-Abelian gauge field theories, as formulated by using classical action functional methods. In Sec. II.B we give the basic formalism for classical Lagrangian statics and illustrate it by briefly considering the case of classical electrostatics. In Sec. II.C we discuss the statics of classical sources in the Abelian Higgs model, in which external source charges are screened. In Sec. II.D we discuss the statics of classical sources using the leading logarithm semiclassical effective action functional for an SU(n) non-Abelian gauge theory, and show analytically that this model describes flux and charge confinement. As a secondary topic we consider non-Abelian gauge field configurations with topological quantum numbers, as exemplified by the axially-symmetric SU(2) topological monopole solutions, the theory of which is discussed in Sec. II.E. In the analyses of Secs. II.C–II.E, we place particular emphasis on identifying special features of the models under study which must be taken into account when solving them numerically.

B. Classical Lagrangian statics

Consider a classical dynamical system described by the action

$$S=\int dtL, \tag{2.1a}$$

$$L=\sum_i{}' p_i\dot{q}_i-H(p,q),\ \dot{q}=\frac{dq}{dt}\ , \tag{2.1b}$$

where q_i and p_i are the canonical coordinates and momenta, where H is the Hamiltonian, and where the prime indicates that those coordinates which have identically vanishing canonical momenta are omitted from the sum. We will be specifically interested in systems for which the equations of motion implied by extremizing S have nontrivial time-independent solutions, and want to find a variational principle for calculating the energy

$$V_{\text{static}}=H(p,q) \tag{2.2}$$

associated with such static solutions. For time-independent solutions, extremizing S is equivalent to extremizing $L(q;\dot{q}=0)$, and so evaluating Eq. (2.1b) at $\dot{q}=0$ gives

$$\begin{aligned}L_{\text{ext}}&=\text{ext}_q L(q;\dot{q}=0)\\&=-H=-V_{\text{static}}\ .\end{aligned} \tag{2.3}$$

Equation (2.3) gives a variational formulation of the problem of calculating V_{static} and is the fundamental equation of classical Lagrangian statics.

As an illustration of Eq. (2.3), let us consider the familiar example of classical electrostatics. The Lagrangian for the Maxwell field coupled to an external static source density j^0 is[1]

$$L=\int d^3x[\tfrac{1}{2}(\mathbf{E}^2-\mathbf{B}^2)-j^0A^0]\ , \tag{2.4}$$

where the fields $\mathbf{E}$ and $\mathbf{B}$ are related to the scalar potential A^0 and the vector potential $\mathbf{A}$ by

$$\mathbf{E}=-\nabla A^0-\dot{\mathbf{A}},\ \mathbf{B}=\nabla\times\mathbf{A}\ . \tag{2.5}$$

Specializing to static solutions with $\dot{\mathbf{A}}=0$, we have

$$L[A^0,\mathbf{A};\dot{\mathbf{A}}=0]=\int d^3x\{\tfrac{1}{2}[(\nabla A^0)^2-(\nabla\times\mathbf{A})^2]-j^0A^0\}\ , \tag{2.6}$$

which is stationary when the potentials satisfy

$$\nabla\cdot\mathbf{E}=-\nabla^2A^0=j^0\ , \tag{2.7a}$$

$$\nabla\times\mathbf{B}=\nabla\times(\nabla\times\mathbf{A})=0\ . \tag{2.7b}$$

If the potentials are required to vanish at infinity, the general solution to Eq. (2.7) is

$$\begin{aligned}A^0(\mathbf{x})&=\int d^3x'\frac{j^0(\mathbf{x}')}{4\pi|\mathbf{x}-\mathbf{x}'|}\ ,\\ \mathbf{A}(\mathbf{x})&=\nabla\Psi(\mathbf{x})\ ,\end{aligned} \tag{2.8}$$

with Ψ an arbitrary gauge function. Substituting Eqs. (2.7) and (2.8) back into Eq. (2.6), we get, after an integration by parts,

$$\begin{aligned}L_{\text{ext}}&=\int d^3x(-\tfrac{1}{2}A^0\nabla^2A^0-j^0A^0)=-\tfrac{1}{2}\int d^3x\,j^0A^0\\&=-\tfrac{1}{2}\int d^3x d^3x'\frac{j^0(\mathbf{x})j^0(\mathbf{x}')}{4\pi|\mathbf{x}-\mathbf{x}'|}=-V_{\text{static}}\ ,\end{aligned} \tag{2.9}$$

[1]Boldface will be used throughout to denote spatial vector indices.

in agreement with the general formula of Eq. (2.3). As a final remark, we note that this example shows that Eq. (2.3) is not a minimum principle; although Eq. (2.6) is stationary at $\mathbf{B}=\nabla\times\mathbf{A}=0$, this value of $\mathbf{B}$ maximizes, rather than minimizes,[2] the Lagrangian L.

C. The Abelian Higgs model

The first, and simplest nonlinear model which we shall discuss is the Abelian Higgs model, coupled to an external source charge density. The fields of this model are an Abelian gauge potential A^μ together with a complex scalar field φ of charge e. The Lagrangian is[3]

$$L=\int d^3x\mathscr{L}\,,\qquad(2.10)$$
$$\mathscr{L}=\tfrac{1}{2}(\mathbf{E}^2-\mathbf{B}^2)-j^0A^0+\left|\left[i\frac{\partial}{\partial t}-eA^0\right]\varphi\right|^2-|(i\nabla+e\mathbf{A})\varphi|^2-\tfrac{1}{2}C(|\varphi|^2-\kappa^2)^2\,,$$

with $\mathbf{E}$ and $\mathbf{B}$ constructed from the potentials $\mathbf{A}$ and A^0 as in Eq. (2.5), and with $|\varphi|^2=\varphi\varphi^*$, where φ^* is the complex conjugate of φ. When specialized to time-independent fields by setting $\dot{\mathbf{A}}$ and $\dot{\varphi}$ to zero, Eq. (2.10) simplifies to

$$\begin{aligned}\mathscr{L}=&\tfrac{1}{2}[(\nabla A^0)^2-(\nabla\times\mathbf{A})^2]-j^0A^0\\&+e^2(A^0)^2|\varphi|^2-|(i\nabla+e\mathbf{A})\varphi|^2\\&-\tfrac{1}{2}C(|\varphi|^2-\kappa^2)^2\,,\end{aligned}\qquad(2.11)$$

which is invariant under the time-independent gauge transformation

$$\begin{aligned}&\varphi\to\varphi e^{i\psi}\,,\\&\mathbf{A}\to\mathbf{A}+e^{-1}\nabla\psi\,.\end{aligned}\qquad(2.12)$$

By a suitable choice of gauge[3] we can make the scalar field φ real, so that Eq. (2.11) becomes

$$\begin{aligned}\mathscr{L}=&\tfrac{1}{2}[(\nabla A^0)^2-(\nabla\times\mathbf{A})^2]-j^0A^0\\&+e^2(A^0)^2\varphi^2-(\nabla\varphi)^2-e^2\mathbf{A}^2\varphi^2-\tfrac{1}{2}C(\varphi^2-\kappa^2)^2\,.\end{aligned}\qquad(2.13)$$

As our final simplification, we note that since Eq. (2.13) has no source term coupled to $\mathbf{A}$, it is stationary with respect to variations of $\mathbf{A}$ around $\mathbf{A}=0$. Hence to calculate V_{static} it suffices to consider the $\mathbf{A}=0$ specialization of Eq. (2.13), giving

$$\begin{aligned}L[A^0,\varphi]=&\int d^3x\{\tfrac{1}{2}(\nabla A^0)^2-j^0A^0+e^2(A^0)^2\varphi^2\\&-(\nabla\varphi)^2-\tfrac{1}{2}C(\varphi^2-\kappa^2)^2\}\,,\\V_{\text{static}}=&-\text{ext}_{A^0,\,\varphi}L[A^0,\varphi]\,.\end{aligned}\qquad(2.14)$$

Varying the action of Eq. (2.14), we get the Euler-Lagrange equations

$$\begin{aligned}&\nabla^2A^0=2e^2A^0\varphi^2-j^0\,,\\&\nabla^2\varphi=-e^2(A^0)^2\varphi+C\varphi(\varphi^2-\kappa^2)\,.\end{aligned}\qquad(2.15)$$

These equations will be solved numerically in Sec. IV.B for a source j^0 describing point charges located symmetrically on the z axis,

$$j^0=Q\delta(x)\delta(y)[\delta(z-a)-\delta(z+a)]\,,\qquad(2.16)$$

for which A^0 is an even function and φ is an odd function of z. A straightforward analysis[3] shows that the leading behavior of A^0 and φ, at infinity and in the neighborhood of the source charges, is given by the following formulas.

At ∞:

$$\begin{aligned}&\varphi\sim\kappa+\frac{\varphi^{(\infty)}}{r}\exp[-r(2C\kappa^2)^{1/2}]\,,\\&A^0\sim A^{0(\infty)}\left[\frac{1}{r_1}-\frac{1}{r_2}\right]\exp[-r(2e^2\kappa^2)^{1/2}]\,.\end{aligned}\qquad(2.17a)$$

At $r_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}\sim0$:

$$\begin{aligned}&\varphi\sim\varphi^{(0)}r^{\lambda}_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]},\quad\lambda=-\tfrac{1}{2}+\left[\tfrac{1}{4}-\left[\frac{eQ}{4\pi}\right]^2\right]^{1/2},\\&A^0\sim(\pm)\left[\frac{Q}{4\pi r_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}}+A^{0(0)}\right],\end{aligned}\qquad(2.17b)$$

with $\varphi^{(\infty)}$, $A^{0(\infty)}$, $\varphi^{(0)}$, $A^{0(0)}$ constants and with

$$\begin{aligned}&r=(x^2+y^2+z^2)^{1/2}\,,\\&r_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}=[x^2+y^2+(z\mp a)^2]^{1/2}\,.\end{aligned}\qquad(2.18)$$

At large distances, the Higgs field φ approaches the constant κ (there is a second solution to the equations with $\varphi\to-\varphi$) and the scalar potential A^0 shows the characteristic exponential decay expected in a Higgs phase, which arises from the screening of the source j^0 by the charged Higgs field. Close to the source charges, the Higgs field becomes infinite as $r^{\lambda}_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}$ with $-\tfrac{1}{2}<\lambda<0$ for

[2]For neutral charge distributions (with $\int d^3j^0=0$) the variational principle $\delta L[A^0,\mathbf{A};\dot{\mathbf{A}}=0]=0$ is minimax: the fields of classical electrostatics minimize L with respect to variations in A^0, while maximizing L with respect to variations in $\mathbf{A}$. A functional which (for neutral charge distributions) is minimized by the fields of classical electrostatics is

$$\tilde{L}[A^0,\mathbf{A}]=\int d^3x\{\tfrac{1}{2}[(\nabla A^0)^2+(\nabla\times\mathbf{A})^2]-j^0A^0\}\,.$$

Functionals of this form can be useful for mathematical purposes [see, e.g., Adler (1981a, 1981b) and Footnote 13 below], but unlike L have no direct physical interpretation.

[3]For a pedagogical discussion of the Abelian Higgs model, see Bernstein (1974). The analysis described in Sec. II.C was carried out by Adler and Pearson (1978, and unpublished); see Appendix B of Adler (1978a).

weak source charges Q satisfying the inequality[4]

$$\left[\frac{eQ}{4\pi}\right]^2 < \frac{1}{4} , \tag{2.19}$$

and the scalar potential has Coulomb-type singularities. The removal of the infinite Coulomb self-energies from the formula for V_{static} will be discussed in detail later on, in Sec. III.E of the text and in Appendix C.

D. Non-Abelian statics in the leading logarithm model

The second nonlinear model which we shall discuss is constructed from an SU(n) non-Abelian gauge theory coupled to an external source charge density. The fields of this model are an SU(n) non-Abelian gauge potential $A^{b\mu}$, with $b=1,\ldots,n^2-1$ the internal symmetry index. Making the conventional rescaling of the gauge potentials by the coupling constant g, the action and Lagrangian for this theory are

$$S=\int L dt,\; L=\int d^3x \mathscr{L} ,$$
$$\mathscr{L}=\frac{1}{2g^2}(\mathbf{E}^a\cdot\mathbf{E}^a-\mathbf{B}^a\cdot\mathbf{B}^a)-j^{a0}A^{a0} , \tag{2.20}$$

with the field-potential relations given now by

$$E^{aj}=-\mathscr{D}_jA^{a0}-\frac{\partial}{\partial t}A^{aj} ,$$
$$B^{aj}=\varepsilon^{jkl}\left[\frac{\partial}{\partial x^k}A^{al}+\tfrac{1}{2}f^{abc}A^{bk}A^{cl}\right] . \tag{2.21}$$

In Eq. (2.21), ε^{jkl} denotes the usual three-index antisymmetric tensor defined by

$$\varepsilon^{jkl}=\varepsilon^{ljk}=\varepsilon^{klj} ,\; \varepsilon^{jkl}=-\varepsilon^{kjl},\; \varepsilon^{123}=1 , \tag{2.22}$$

f^{abc} are the SU(n) group structure constants [for SU(2), $f^{abc}=\varepsilon^{abc}$], and $\mathscr{D}_j$ is the covariant derivative defined (for arbitrary w^a) by

$$\mathscr{D}_jw^a=\frac{\partial}{\partial x^j}w^a+f^{abc}A^{bj}w^c . \tag{2.23}$$

Static (and nonstatic) extrema of the action of Eq. (2.20) have been extensively discussed in the literature,[5] and can be found numerically by the same algorithm which we use later on to solve the topological monopole model. Hence instead of basing a numerical example on Eq. (2.20), we consider instead the much more interesting model in which L is replaced by an effective action L_{eff}, which (for slowly varying fields) incorporates the effect of radiative corrections to leading logarithm order, while keeping j^{a0} a classical[6] source. Both explicit one-loop calculations for the special case of constant field-strengths,[7] and more general renormalization-group arguments,[8] show that the action L_{eff} is obtained by replacing the coupling constant g^2 in Eq. (2.20) by a field-strength-dependent "running" coupling constant $g^2(\mathscr{F})$,

[4]The fact that λ becomes complex for large Q is an indication that, for large Q, pair production is important and that a field-theoretic discussion is needed. In a full field-theory treatment of the Abelian Higgs model, Eq. (2.14) is replaced by

$$V_{\text{static}}=-\text{ext}_{A^0,\varphi}L_{\text{eff}}[A^0,\varphi] ,$$

with L_{eff} an effective action functional which includes the effects of virtual quanta. When radiative corrections are ignored, L_{eff} reduces to the Lagrangian of Eq. (2.14).

[5]For an exhaustive survey, see Actor (1979).

[6]Two types of approximation schemes have been discussed in the literature for reducing SU(n) quantum chromodynamics with quantized source charges to classical source charge models. For methods involving a direct replacement of the SU(3) color charges by quasi-Abelian effective charges which respect the triality selection rules for color singlet states, see Mandula (1976) and Adler (1982). For methods involving a study of the algebraic properties of the SU(n) color charges, see Khriplovich (1978); Adler (1978b); Giles and McLerran (1978); Cvitanović, Gonsalves, and Neville (1978); Rittenberg and Wyler (1978); Lee (1979); Adler (1979); Lee (1980); Adler (1980); Bender, Gromes, and Rothe (1980); Adler (1981a); Milton, Palmer, and Pinsky (1982); and Milton, Wilcox, Palmer, and Pinsky (1982).

[7]The one-loop Yang-Mills effective action functional for constant field strengths has been calculated by a number of authors. See, for an early calculation, Batalin, Matinyan, and Savvidy (1977), and for recent discussions and references, Schanbacher (1982) and Anishetty (1982). Methods for constructing gauge-invariant effective action functionals beyond one-loop order have been given by 't Hooft (1975a), DeWitt (1981), Boulware (1981), and Abbott (1981).

[8]Matinyan and Savvidy (1978) and Pagels and Tomboulis (1978) have shown how the structure of the renormalization-group improved effective action can be inferred from the conformal trace anomaly. Renormalization-group arguments give an expression for $\mathscr{L}(\mathscr{F})$ of the form

$$\mathscr{L}(\mathscr{F})=\tfrac{1}{8}b_0\mathscr{F}\log(\mathscr{F}/e\kappa^2)\left[1+\frac{8b_1}{b_0^2}\frac{\log\log(\mathscr{F}/e\kappa^2)}{\log(\mathscr{F}/e\kappa^2)}\right]+O(\mathscr{F}) ,$$

with b_0 and b_1 the usual β-function coefficients defined in one- and two-loop orders. Adler (1981b) has argued that this expression may give the leading two terms in the effective action for *weak* fields $|\mathscr{F}/e\kappa^2|\ll 1$ as well as for strong fields $|\mathscr{F}/e\kappa^2|\gg 1$, because the magnitude of the running coupling constant of Eq. (2.24) is small in both regions. This argument suggests that, very generally, the effective dielectric constant changes sign between the strong and weak field regions, which is the essential feature responsible for confinement in the leading logarithm model.

The corrections of order $\mathscr{F}$ are not determined by renormalization-group arguments and in general are highly nonlocal (i.e., they depend on derivatives of the field strengths). Adler (1983) gives arguments indicating that the nonlocal terms in L_{eff} become important in the ultraviolet (short-distance) limit, but are unimportant relative to the local terms in the infrared (long-distance, or confining) limit. The order-$\mathscr{F}$ terms also can have imaginary parts; for example, if κ^2 in Eq. (2.28) is replaced by $-\kappa^2$, an additional imaginary term appears in $\mathscr{L}(\mathscr{F})$ at the order-$\mathscr{F}$ level. Hence even the sign of $\mathscr{F}$ at the extremum of $\mathscr{L}$ cannot be determined by a renormalization-group argument.

$$g^2(\mathcal{F})=\frac{g^2}{1+\frac{1}{4}b_0g^2\log(\mathcal{F}/\mu^4)}\,, \tag{2.24}$$

with $\mathcal{F}$ the field-strength invariant

$$\mathcal{F}=\mathbf{E}^a\cdot\mathbf{E}^a-\mathbf{B}^a\cdot\mathbf{B}^a, \tag{2.25}$$

which appears in the classical action. The constant b_0 in Eq. (2.24) is the asymptotic freedom constant[9]

$$b_0=\frac{1}{8\pi^2}\frac{11}{3}C_2[\mathrm{SU}(n)]=\frac{1}{8\pi^2}\frac{11}{3}n\,, \tag{2.26}$$

while the mass μ is the renormalization point and g^2 is the value of the running coupling constant at $\mathcal{F}=\mu^4$. Combining Eqs. (2.20)–(2.25) and defining the one-loop renormalization-group-invariant parameter

$$\kappa^2=\frac{\mu^4}{e}e^{-4/(b_0g^2)}\,, \tag{2.27}$$

we get the effective action for the leading logarithm model,[8]

$$L_{\mathrm{eff}}=\int d^3x\,\mathcal{L}_{\mathrm{eff}}=\int d^3x[\mathcal{L}(\mathcal{F})-j^{a0}A^{a0}]\,, \tag{2.28a}$$

$$\mathcal{L}(\mathcal{F})=\tfrac{1}{8}b_0\mathcal{F}\log(\mathcal{F}/e\kappa^2)\,. \tag{2.28b}$$

When specialized to time-independent fields,[10] Eq. (2.21) for E^{aj} becomes

$$E^{aj}=-\mathcal{D}_jA^{a0}\,, \tag{2.29}$$

and Eq. (2.3) for V_{static} yields the variational problem

$$V_{\mathrm{static}}=-\mathrm{ext}_{A^{a0},A^{aj}}\{L_{\mathrm{eff}}[A^{a0},A^{aj}]\}\,. \tag{2.30}$$

The Euler-Lagrange equations for Eq. (2.30) are

$$\mathcal{D}_j(\varepsilon E^{aj})=j^{a0}\,, \tag{2.31a}$$

$$\varepsilon^{kjm}\mathcal{D}_j(\varepsilon B^{am})=-f^{abc}A^{b0}\varepsilon E^{ck}\,, \tag{2.31b}$$

where we have introduced a field-strength-dependent effective dielectric constant ε defined by

$$\varepsilon=\frac{\partial\mathcal{L}(\mathcal{F})}{\partial(\frac{1}{2}\mathcal{F})}=\tfrac{1}{4}b_0\log(\mathcal{F}/\kappa^2)\,. \tag{2.32}$$

Applying a covariant derivative $\mathcal{D}_k$ to Eq. (2.31b) gives the equation

$$\tfrac{1}{2}\varepsilon^{kjm}[\mathcal{D}_k,\mathcal{D}_j](\varepsilon B^{am})=-f^{abc}(\mathcal{D}_kA^{b0})\varepsilon E^{ck}-f^{abc}A^{b0}\mathcal{D}_k(\varepsilon E^{ck})\,. \tag{2.33}$$

Using the easily verified identity (which holds for arbitrary w^a),

$$[\mathcal{D}_k,\,\mathcal{D}_j]w^a=\varepsilon^{kjl}f^{abc}B^{bl}w^c\,, \tag{2.34}$$

we find that the left-hand side of Eq. (2.33) is

$$\tfrac{1}{2}\varepsilon^{kjm}\varepsilon^{kjl}f^{abc}B^{bl}\varepsilon B^{cm}=0\,, \tag{2.35a}$$

while on substituting Eq. (2.29) the first term on the right-hand side of Eq. (2.33) becomes

$$f^{abc}\varepsilon E^{bk}E^{ck}=0\,. \tag{2.35b}$$

Hence the second term on the right-hand side of Eq. (2.33) must also vanish. After substitution of Eq. (2.31a), this gives the constraint

$$f^{abc}A^{b0}j^{c0}=0\,, \tag{2.36}$$

which states that to get a static solution, the scalar potential and the source charge density must locally lie in commuting directions in internal symmetry space.

In particular, for a source density j^{c0} describing a particle with effective classical charge Q^c at $\mathbf{x}_1$ and an antiparticle with effective classical charge $\bar{Q}^c$ at $\mathbf{x}_2$,

$$j^{c0}=Q^c\delta^3(\mathbf{x}-\mathbf{x}_1)+\bar{Q}^c\delta^3(\mathbf{x}-\mathbf{x}_2)\,, \tag{2.37}$$

the constraint of Eq. (2.36) becomes

$$\begin{aligned}&f^{abc}A^{b0}(\mathbf{x}_1)Q^c=0\,,\\&f^{abc}A^{b0}(\mathbf{x}_2)\bar{Q}^c=0\,.\end{aligned} \tag{2.38}$$

By making a suitable time-independent gauge transformation, we can align Q^c and $\bar{Q}^c$ to lie in antiparallel directions in internal symmetry space,

$$Q^c=\hat{q}^cQ\,,\ \bar{Q}^c=-\hat{q}^cQ\,, \tag{2.39}$$

with $\hat{q}$ a fixed internal symmetry unit vector. The constraints of Eq. (2.38) can then be satisfied by making the quasi-Abelian ansatz

$$\begin{aligned}&A^{a0}=\hat{q}^aA^0\,,\\&A^{aj}=\hat{q}^aA^j,\ \ (A^1,A^2,A^3)\equiv\mathbf{A}\,,\end{aligned} \tag{2.40}$$

which describes a conserved electric flux of magnitude Q running between the two point sources, as is appropriate to a model for the quark-antiquark confinement problem.[11] For the potentials of Eq. (2.40), the field-potential relations of Eq. (2.21) become

---

[9]In SU(3) quantum chromodynamics (QCD) with N_f massless fermion flavors, Eq. (2.26) for b_0 becomes

$$b_0=\frac{1}{8\pi^2}(11-\tfrac{2}{3}N_f)\,.$$

[10]Since the physically relevant extrema of the effective action are the *mean* potentials induced when a source j^{a0} is added to the standard functional integration quantization formalism [see, for example, Abers and Lee (1973)], they must be time independent when the source is time independent.

[11]The quasi-Abelian ansatz of Eq. (2.40) excludes "charge-screening" solutions of the type discussed by Sikivie and Weiss (1978, 1979), Kiskis (1980a), Jackiw and Rossi (1980), and Hilf and Polley (1981). Such solutions may be relevant as models for the behavior of an SU(n) gauge field with *adjoint* representation sources. *Fundamental* representation sources, such as quarks and antiquarks, cannot be screened by the gauge gluon field.

$$E^{aj}=\hat{q}^a E^j,\ (E^1,E^2,E^3)\equiv \mathbf{E}=-\nabla A^0\,,$$
$$B^{aj}=\hat{q}^a B^j,\ (B^1,B^2,B^3)\equiv \mathbf{B}=\nabla\times\mathbf{A}\,. \qquad (2.41)$$

The internal symmetry structure of the problem can now be completely factored away. Equations (2.30) and (2.28) simplify to

$$\begin{aligned} V_{\rm static}&=-{\rm ext}_{A^0,A^j}\{L_{\rm eff}[A^0,A^j]\}\,,\\ L_{\rm eff}&=\int d^3x[\mathscr{L}(\mathscr{F})-j^0A^0]\,,\\ \mathscr{F}&=(\nabla A^0)^2-(\nabla\times\mathbf{A})^2\,,\\ j^0&=Q\delta(x)\delta(y)[\delta(z-a)-\delta(z+a)]\,, \end{aligned}\qquad (2.42)$$

where we have again located the source charges symmetrically on the z axis, and the Euler-Lagrange equations of Eq. (2.31) become

$$\nabla\cdot(\varepsilon\mathbf{E})=j^0\,, \qquad (2.43a)$$

$$\nabla\times(\varepsilon\mathbf{B})=0\,, \qquad (2.43b)$$

with the dependence on $\mathscr{F}$ of $\mathscr{L}$ and ε given by Eq. (2.28b) and Eq. (2.32), respectively. We have thus reduced our model to a problem in nonlinear Abelian electrostatics.[12,13]

As in the discussion of classical electrostatics in Sec. II.B, the extremum over the vector potential in Eq. (2.42) can be carried out by inspection. From Eq. (2.43b) we get

$$\begin{aligned} 0&=\int d^3x\mathbf{A}\cdot\nabla\times(\varepsilon\mathbf{B})\\ &=\int d^3x\varepsilon\mathbf{B}^2-\int_{\text{surface at }\infty} d\mathbf{S}\cdot\varepsilon(\mathbf{A}\times\mathbf{B})\,, \end{aligned}\qquad (2.44)$$

and so if we restrict ourselves to solutions with a vanishing surface integral at infinity, we must have

$$\varepsilon\mathbf{B}^2=0\,, \qquad (2.45a)$$

giving three branches (Ia, Ib, and II, respectively),

$$\begin{aligned} &\mathbf{B}=0\,,\ \mathbf{E}^2>\kappa^2\,,\\ &\mathbf{B}=0\,,\ \mathbf{E}^2<\kappa^2\,,\\ &\varepsilon=0\ \Longrightarrow\ \mathbf{B}^2=\mathbf{E}^2-\kappa^2\,. \end{aligned}\qquad (2.45b)$$

In the strong-field region near the source charges, the asymptotic freedom of non-Abelian gauge theories requires that the solution of Eqs. (2.43) approach a Coulomb-type solution with $\mathbf{E}$ large and $\mathbf{B}$ vanishing; together with continuity, this implies that a finite domain containing the source charges lies on branch Ia. Specializing the analysis, for the time being, to this branch, we set $\mathbf{B}=\mathbf{A}=0$ and rewrite Eqs. (2.41) and (2.43a) as

$$\begin{aligned} &\nabla\cdot\mathbf{D}=j^0\,,\ \nabla\times\mathbf{E}=0\,,\\ &\mathbf{D}=\varepsilon(E)\mathbf{E}\,,\\ &\varepsilon(E)=\tfrac{1}{4}b_0\log(E^2/\kappa^2)\,,\ E=|\mathbf{E}|\,. \end{aligned}\qquad (2.46)$$

A graph of the nonlinear dielectric constant $\varepsilon(E)$, showing its intersection (for $\mathbf{B}=0$) with the three branches of Eq. (2.45b), is shown in Fig. 1.

Equations (2.46) are the basic statement of the problem which will be studied numerically in Sec. IV.C. In order to get a tractable numerical method, it is necessary (for reasons discussed in Appendix A) to rewrite the equations in manifestly flux-conserving form. To exploit the axial symmetry of the problem, let us work henceforth in cylindrical coordinates ρ,z,ϕ defined by

$$\rho=(x^2+y^2)^{1/2},\ \ \phi=\tan^{-1}(y/x)\,, \qquad (2.47)$$

in which the coordinates of the point sources of Eq. (2.42) are $\rho=0$, $z=\pm a$. We then note that $\mathbf{D}$ can be parametrized in terms of a single scalar function $\Phi(\rho,z)$ by writing[14]

$$\mathbf{D}=-\frac{1}{2\pi}\nabla\phi\times\nabla\Phi=-\frac{\hat{\phi}}{2\pi\rho}\times\nabla\Phi=\nabla\times\left[\frac{\hat{\phi}}{2\pi\rho}\Phi\right]. \qquad (2.48)$$

The representation of Eq. (2.48) automatically satisfies $\nabla\cdot\mathbf{D}=0$ at points off the axis, and at points on the axis where Φ is sufficiently smooth. The physical interpretation of Φ follows from calculating the total flux through a surface of revolution S (with element of area $d\mathbf{A}$) bounded by a circle C of radius ρ (with element of arc length $d\mathbf{l}=dl\hat{\phi}$), as shown in Fig. 2. We get

$$\begin{aligned} \text{flux through } S&=\int_S d\mathbf{A}\cdot\mathbf{D}=\int_S d\mathbf{A}\cdot\nabla\times\left[\frac{\hat{\phi}}{2\pi\rho}\Phi\right]\\ &=\int_C d\mathbf{l}\cdot\left[\frac{\hat{\phi}}{2\pi\rho}\Phi\right]=\Phi\,, \end{aligned}\qquad (2.49)$$

showing that Φ is simply the flux through S. If we draw the surface S so that it always intersects the z axis on the segment $z>a$, as shown in Fig. 2, the flux function Φ as-

[12]Dielectric models for confinement in QCD have been discussed in a qualitative way by a number of authors; see, for example, Kogut and Susskind (1974), 't Hooft (1975b), Pagels and Tomboulis (1978), Friedberg and Lee (1978), Callan, Dashen, and Gross (1979), and Nambu (1981).

[13]Pagels and Tomboulis (1978) and Mills (1979) showed that when a single isolated charge is present, the leading logarithm model gives a linearly divergent infrared energy. A proof that the model of Eq. (2.42) gives a linear static potential for large source separations was first given by Adler (1981a), using the related minimum principle in which $\mathscr{F}$ is replaced by $(\nabla A^0)^2+(\nabla\times\mathbf{A})^2$. (See the comments in Footnote 2 above.)

[14]The flux function formulation was introduced in Adler (1981b). The analysis of the characteristic form of the flux function equation, and its numerical solution, were given by Adler and Piran (1982a).

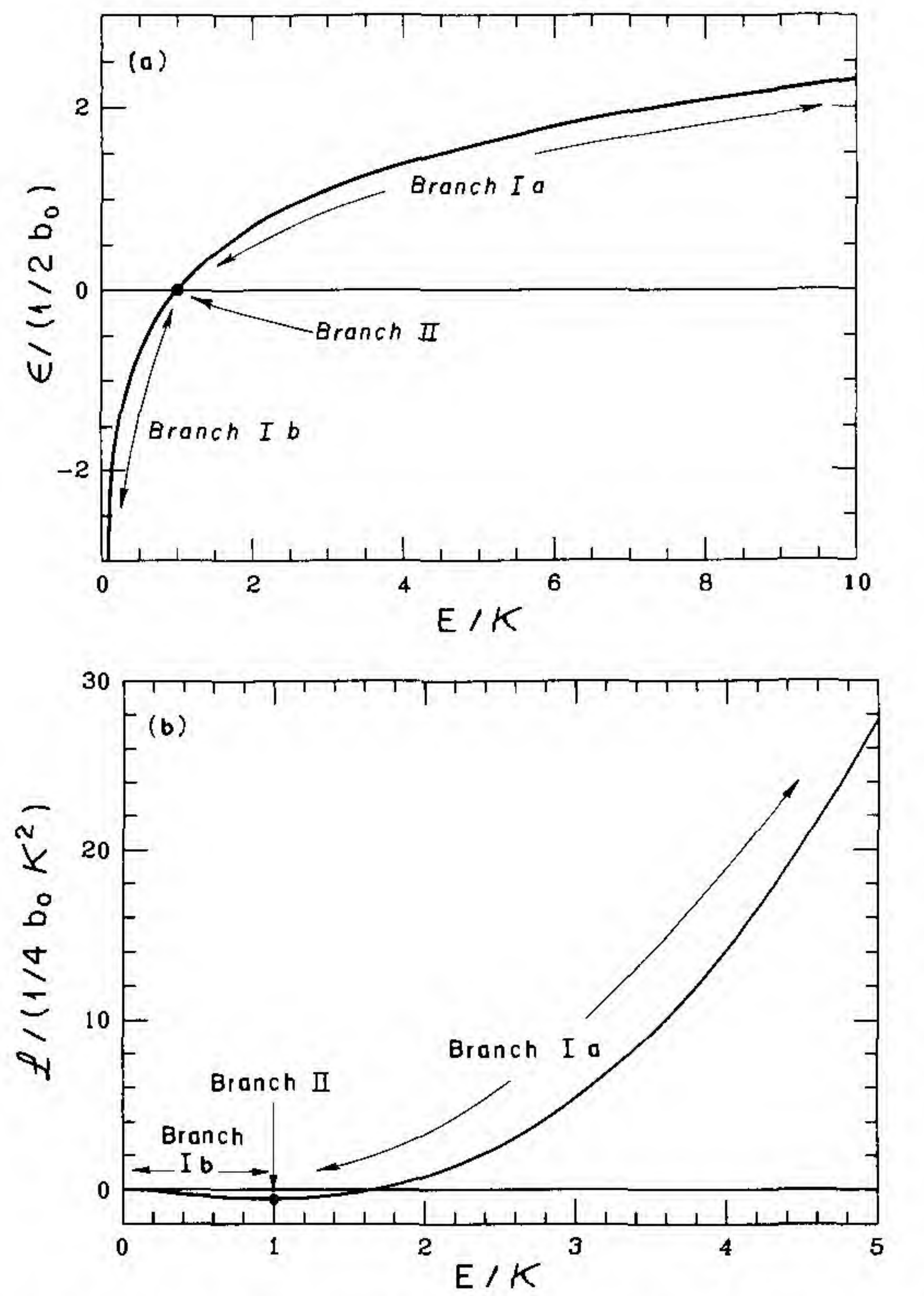


FIG. 1. (a) Plot of $\varepsilon(E)$ of Eq. (2.46), showing its intersection (for $\mathbf{B}=0$) with the three branches of Eq. (2.45b). (b) Corresponding plot of $\mathscr{L}(\mathscr{F}=E^2)$ of Eq. (2.28b).

sumes the following boundary values on the axis of rotation and at infinity:

$$\begin{aligned} &\Phi=0,\ \rho=0,\ |z|>a\ , \\ &\Phi=Q,\ \rho=0,\ |z|<a\ , \\ &\Phi\to 0 \text{ as } \rho^2+z^2\to\infty\ . \end{aligned} \tag{2.50}$$

To verify these, we note that Φ is an even function of z, and that on the segment $\rho=0$, $z>a$, the surface S degenerates to a point and intercepts no flux. Similarly, on the segment $\rho=0$, $|z|<a$, the surface S intercepts all of the flux in a positive sense, as illustrated in Fig. 3. The requirement that Φ should vanish on the sphere at infinity

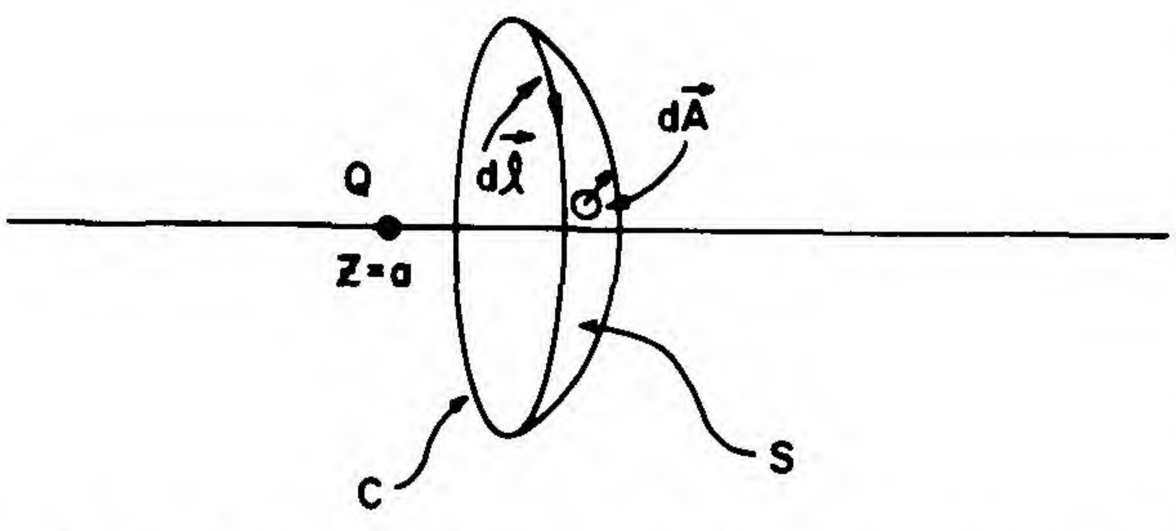


FIG. 2. Surface of revolution S with circular boundary C used in Eq. (2.49) to evaluate the flux function Φ.

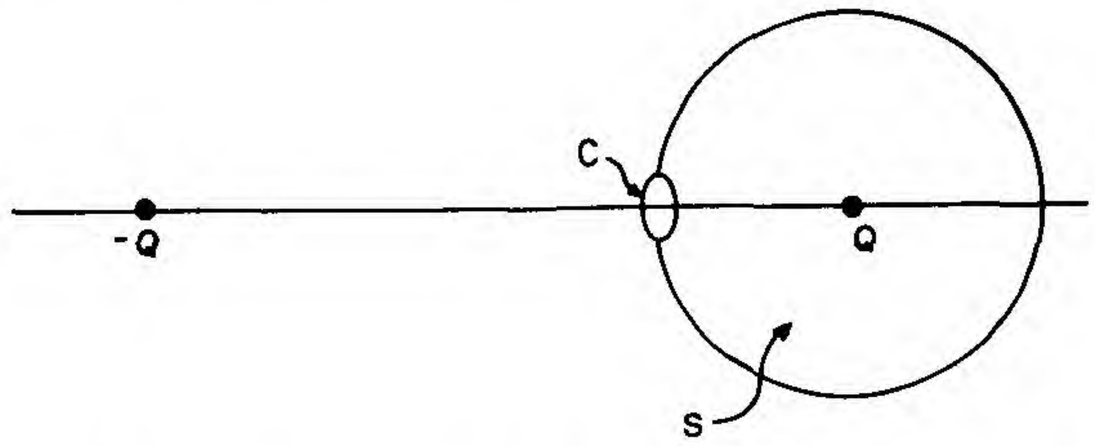


FIG. 3. Surface S (bounded by an infinitesimal circle C) used to evaluate Φ on the axis at $|z|<a$.

ensures that no additional flux sources or sinks lie at spatial infinity.

The dynamical equation for Φ is obtained from

$$\nabla\times\mathbf{E}=\nabla\times\left[\frac{\mathbf{D}}{\varepsilon}\right]=0\ . \tag{2.51}$$

Defining $D=|\mathbf{D}|$, we can algebraically invert the constitutive equation $D=\varepsilon(E)E$ to get

$$\varepsilon(E(D))\equiv\varepsilon[D]\ , \tag{2.52}$$

so that Eq. (2.51) becomes a differential equation for $\mathbf{D}$,

$$\nabla\times\left[\frac{\mathbf{D}}{\varepsilon[D]}\right]=0\ . \tag{2.53}$$

This equation can be rewritten by using the vector identity

$$\nabla\cdot(\mathbf{V}_1\times\mathbf{V}_2)=\mathbf{V}_2\cdot(\nabla\times\mathbf{V}_1)-\mathbf{V}_1\cdot(\nabla\times\mathbf{V}_2)\ , \tag{2.54}$$

with

$$\begin{aligned} &\mathbf{V}_1=\frac{\mathbf{D}}{\varepsilon}\ ,\quad \mathbf{V}_2=\frac{\hat{\phi}}{\rho}\ , \\ &\nabla\times\mathbf{V}_1=0=\nabla\times\mathbf{V}_2\ , \end{aligned} \tag{2.55}$$

which when simplified by using $\hat{\phi}\cdot\nabla\Phi=0$ gives

$$\begin{aligned} &\nabla\cdot[\sigma(\rho,|\nabla\Phi|)\nabla\Phi]=0\ , \\ &\sigma(\rho,|\nabla\Phi|)\equiv\frac{1}{\rho^2\varepsilon[D]}\ ,\quad D=\frac{|\nabla\Phi|}{2\pi\rho}\ . \end{aligned} \tag{2.56}$$

Equation (2.56) is the formulation of the leading logarithm model which will be studied numerically in Sec. IV.C. As a check, we note that in the case of classical electrostatics, where $\varepsilon[D]\equiv 1$, Eq. (2.56) and the boundary conditions of Eq. (2.50) are satisfied by

$$\begin{aligned} &\Phi=\tfrac{1}{2}Q(\cos\vartheta_2-\cos\vartheta_1)\ , \\ &\vartheta_1=\tan^{-1}\left[\frac{\rho}{z-a}\right],\ \vartheta_2=\tan^{-1}\left[\frac{\rho}{z+a}\right],\ 0\le\vartheta_{1,2}<\pi\ , \end{aligned} \tag{2.57}$$

which, when substituted into Eq. (2.48), gives the expected result

$$\mathbf{D}=\frac{Q\hat{r}_1}{4\pi r_1^2}-\frac{Q\hat{r}_2}{4\pi r_2^2}\,,$$
$$\hat{r}_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}=(\mathbf{x}-\mathbf{x}_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]})/r_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}\,, \tag{2.58}$$
$$r_{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]}=[x^2+y^2+(z\mp a)^2]^{1/2}\,.$$

In the case where $\varepsilon(E)$ is given by Eq. (2.46), the coefficient function σ is given on branch Ia by

$$\sigma(\rho,|\nabla\Phi|)=\frac{2\pi\kappa}{\rho|\nabla\Phi|}f\left[\frac{|\nabla\Phi|}{\pi b_0\kappa\rho}\right]\,, \tag{2.59}$$

with $f(w)$ implicitly defined by the transcendental equation

$$w=f\log f\,,\quad f\geq 1\,. \tag{2.60}$$

For small w and large w, the behavior of $f(w)$ is given by

$$f=1+w-\tfrac{1}{2}w^2+O(w^3)\,,\quad |w|\ll 1\,, \tag{2.61a}$$

$$f=\frac{w}{\log w}\left\{1+\frac{\log\log w}{\log w}+O\left[\left[\frac{\log\log w}{\log w}\right]^2\right]\right\},\quad w\gg 1\,, \tag{2.61b}$$

as shown in Fig. 4(a), giving σ the behavior graphed in Fig. 4(b). The fact that σ becomes infinite as $f=E/\kappa$ approaches 1 from above (i.e., as $w\propto D$ approaches 0) means that a solution which is initially on branch Ia can approach branch II only as a degenerate limit and cannot cross back and forth between branch Ia and branch Ib. In the vicinity of the source charge at $\rho=0$, $z=a$, Eqs. (2.50), (2.56), and (2.59)–(2.61) can be integrated to give the leading behavior[15]

$$\begin{aligned}
\mathbf{D}&=\frac{Q\hat{r}_1}{4\pi r_1^2}+O(1)\,,\\
\mathbf{E}&=\hat{r}_1\kappa f\left[\frac{Q}{2\pi\kappa b_0 r_1^2}\right]+O(1)\,,\\
A^0&=\kappa\int_{r_1}^{\text{const}}dr_1'f\left[\frac{Q}{2\pi\kappa b_0(r_1')^2}\right]+O(r_1)\,,\\
\Phi&=Q\tfrac{1}{2}(1-\cos\vartheta_1)+O(r_1^2\sin^2\vartheta_1)\,,
\end{aligned} \tag{2.62}$$

where the structure of the subdominant terms $O(\ )$ has been indicated up to powers of $\log r_1$. The corresponding behavior at $\rho=0$, $z=-a$ is obtained by reflecting the formulas of Eq. (2.62) in the $z=0$ plane.

Once Φ has been determined by solving the boundary value problem formulated above, the static potential can be calculated by substituting $\nabla\cdot\mathbf{D}=j^0$ into Eq. (2.42) and integrating by parts. This gives

$$V_{\text{static}}=\int d^3x\{ED-\mathscr{L}[E(D)^2]+\mathscr{L}[E(0)^2=\kappa^2]\}\,, \tag{2.63}$$

where an infinite constant $\int d^3x\mathscr{L}(\kappa^2)$ has been added to Eq. (2.63) to render the integral convergent at spatial infinity. A little algebra shows that Eq. (2.63) is equivalent to the following computationally convenient formula:

$$\begin{aligned}
V_{\text{static}}&=\int d^3x\tfrac{1}{2}\sigma\left[\frac{\nabla\Phi}{2\pi}\right]^2(1+\xi)\,,\\
\xi&=(f^2-1)/(2fw)\,,
\end{aligned} \tag{2.64}$$

with σ, f, and w as defined above. A second useful expression for V_{static} is obtained by using the identity

$$\begin{aligned}
\mathscr{L}[E(D)^2]-\mathscr{L}[E(0)^2]&=\int_{E(0)}^{E(D)}dE'\frac{\partial\mathscr{L}(E'^2)}{\partial E'}\\
&=\int_{E(0)}^{E(D)}dE'D(E')\\
&=ED-\int_0^D dD'E(D')\,,
\end{aligned} \tag{2.65}$$

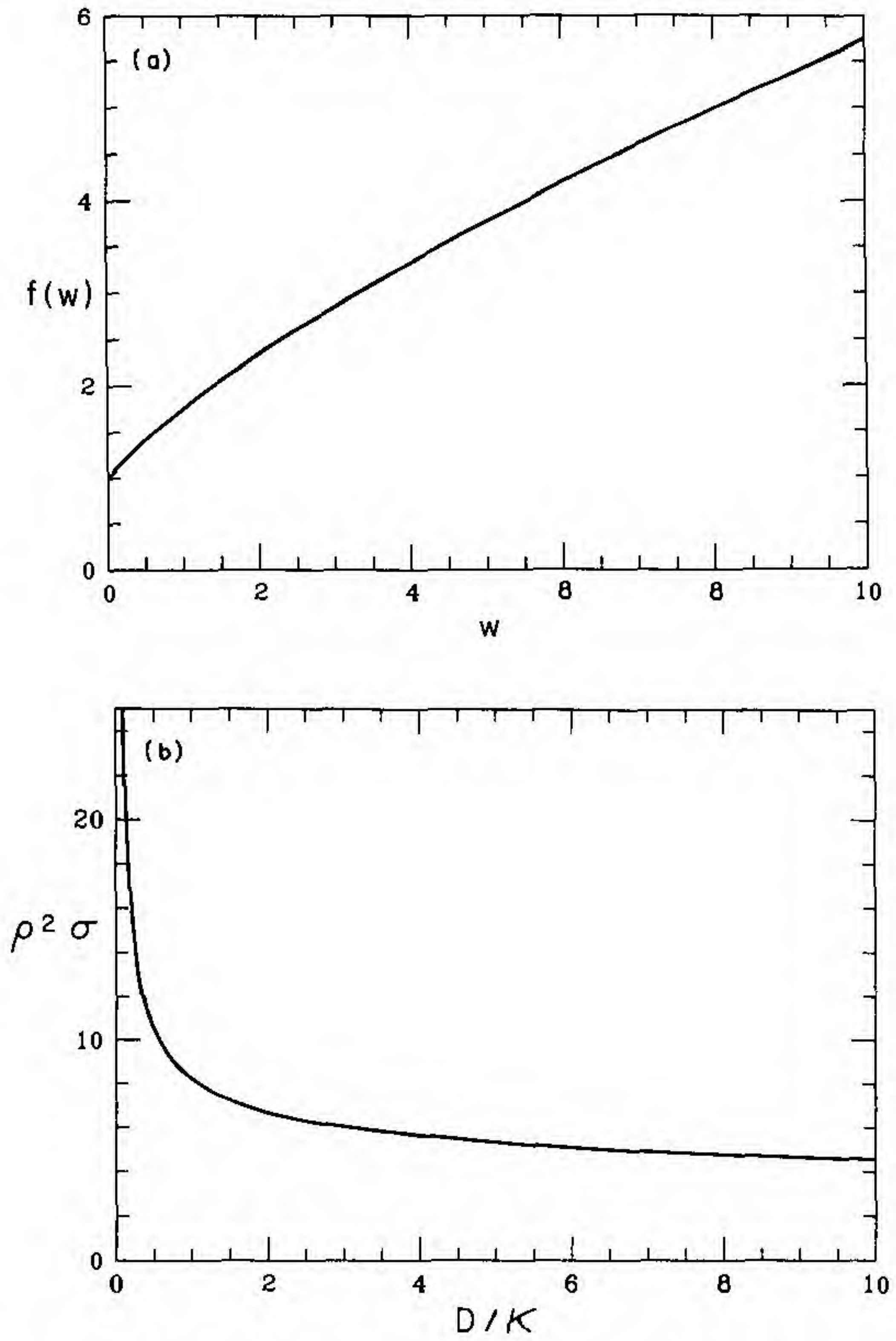


FIG. 4. (a) Plot of the function $f(w)$ defined in Eq. (2.60). (b) Plot of $\rho^2\sigma$, defined in Eqs. (2.59)–(2.61), as a function of D/κ.

[15]Since $Q\frac{1}{2}(1-\cos\vartheta_1)$ exactly satisfies the boundary conditions of Eq. (2.50) around $z=a$, the leading subdominant term in Φ must vanish at $\vartheta_1=0,\vartheta_1=\pi$. This boundary condition eliminates a possible term in Φ behaving as $O[r_1(a+b\cos\vartheta_1)]$, giving the structure shown in Eq. (2.62).

which when substituted into Eq. (2.63) gives the familiar formula

$$V_{\text{static}} = \int d^3x \mathscr{H} , \tag{2.66a}$$

with $\mathscr{H}$ the field energy density

$$\mathscr{H} = \int_0^D dD' E(D') . \tag{2.66b}$$

Since we have noted above (and will see in greater detail below) that the region of support of D is confined to branch Ia, where $E(D) \geq \kappa$, Eq. (2.66) gives the inequality

$$V_{\text{static}} \geq \kappa \int d^3x D . \tag{2.67}$$

To turn Eq. (2.67) into a useful lower bound on the large distance behavior of V_{static}, we must exclude the infinite Coulomb self-energies. This is most easily done by excluding from the x integration small spheres of radius ε around the source charges, motivating the definitions

$$\begin{aligned} &\Omega = \text{domain}\{ \, |\mathbf{x}-\mathbf{x}_1| \geq \varepsilon \, , \, |\mathbf{x}-\mathbf{x}_2| \geq \varepsilon \} \, , \\ &V_{\text{static}}^{(\varepsilon)} = \int_\Omega d^3x \mathscr{H}(x) \geq \kappa \int_\Omega d^3x D \, . \end{aligned} \tag{2.68}$$

When we write $d^3x = dl dA$, with l the length along and dA the element of area perpendicular to the flux lines of D, Eq. (2.67) then yields the lower bound[16]

$$\begin{aligned} &V_{\text{static}}(R) \geq \kappa \int_\Omega dAD \int dl \geq \kappa Q l_{\min} = \kappa Q (R - 2\varepsilon) \, , \\ &R = |\mathbf{x}_1 - \mathbf{x}_2| = 2a \, , \end{aligned} \tag{2.69}$$

showing that V_{static} increases at least linearly for large source separations. A more detailed discussion of the removal of the Coulomb self-energies from the formula for V_{static} is given in Sec. III.E below.

A great deal of insight into the behavior of Eq. (2.56) is obtained by putting it into the standard form for a second-order, quasilinear differential equation,

$$a^{kl}(x,\Phi,\nabla\Phi)\partial_k \partial_l \Phi + c(x,\Phi,\nabla\Phi) = 0 , \tag{2.70}$$

and analyzing the structure of its characteristics. Defining the inward directed unit normal $\hat{n}$ and the corresponding normal derivative ∂_n,

$$\hat{n} = \frac{\nabla\Phi}{|\nabla\Phi|} , \quad \partial_n = \hat{n}\cdot\nabla , \tag{2.71}$$

we can see through a straightforward calculation given in Appendix A that Eq. (2.56) is equivalent to

$$[(\partial_\rho^2 + \partial_z^2 - \partial_n^2) + \alpha\partial_n^2]\Phi - \alpha\rho^{-1}\partial_\rho\Phi = 0 . \tag{2.72}$$

The coefficient α is given by

$$\begin{aligned} &\alpha = 1 + \frac{\partial\log\sigma}{\partial\log|\nabla\Phi|} = \frac{d(\log f)}{d(\log w)} = \frac{wf'(w)}{f(w)} = \frac{w}{w+f(w)} , \\ &w = \frac{\partial_n\Phi}{\pi b_0 \kappa \rho} = \frac{2D}{b_0\kappa} ; \end{aligned} \tag{2.73}$$

[16]The flux estimate of Eq. (2.69) is due to 't Hooft (1975b).

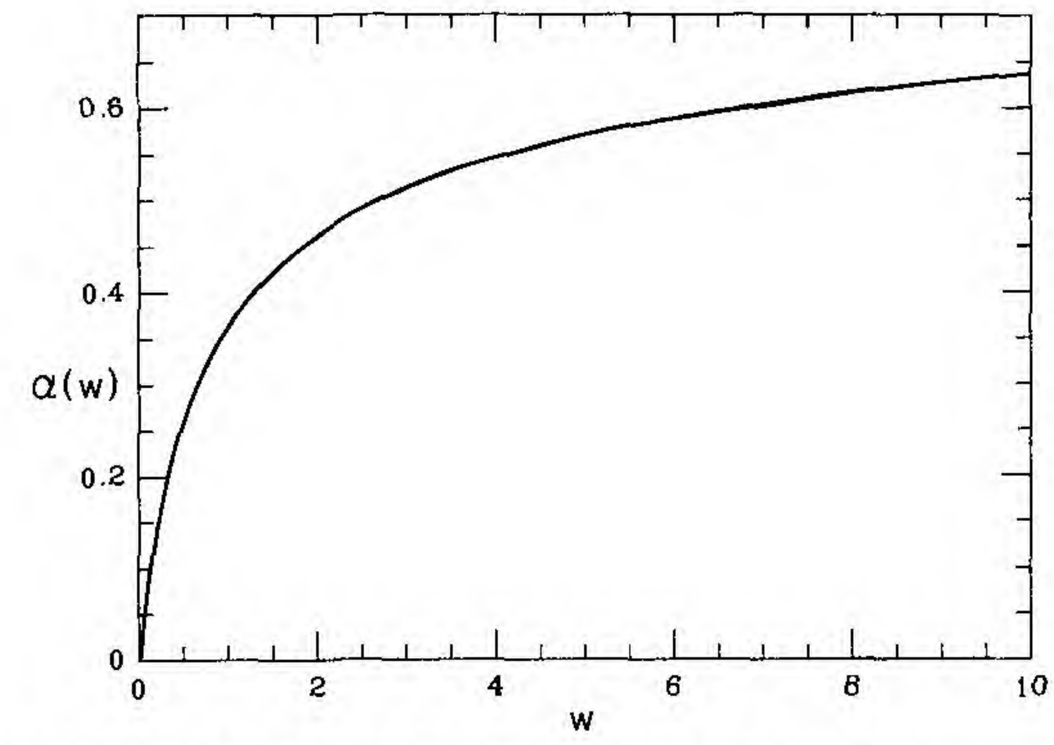


FIG. 5. Plot of the function $\alpha(w)$ defined in Eq. (2.73).

the function $\alpha(w)$ is graphed in Fig. 5 and [from Eq. (2.61)] has the following approximate forms for small and large w:

$$\begin{aligned} &\alpha = w + O(w^2) \, , \quad |w| \ll 1 \, , \\ &\alpha = 1 - \frac{1}{\log w} + O\left[\frac{\log\log w}{(\log w)^2}\right] , \quad w \gg 1 \, . \end{aligned} \tag{2.74}$$

From Eqs. (2.73) and (2.74) and Fig. 5 we see that α lies between 0 and 1 for $D > 0$, but vanishes when $D = 0$. Hence Eq. (2.72) is of degenerating elliptic type,[17] and has a real characteristic at a surface of constant Φ, where $|\nabla\Phi| = 0$. The second normal derivative $\partial_n^2\Phi$ is discontinuous across this characteristic, which acts as a free boundary, dividing space into two causally disconnected regions. From the boundary condition of Eq. (2.50), we learn that the exterior of the free boundary is completely surrounded by surfaces on which $\Phi = 0$. Hence $\Phi \equiv 0$ outside the characteristic, giving the solution the qualitative form graphed in Fig. 6. In the vicinity of a point B on the free boundary where $\rho = \rho_B$ and where the radius of curvature of the free boundary is R_B, a simple analysis given in Appendix A shows that Φ has the leading behavior

$$\Phi \approx \frac{1}{2}\frac{\pi b_0 \kappa \rho_B}{R_B}\left[\left[n - \frac{1}{2}\frac{l^2}{R_B}\right]^2 + O(n^2, l^4)\right] , \tag{2.75}$$

with n and l normal and tangential Cartesian coordinates at B (see Fig. 7). Since Φ is increasing towards the interior, we must have $R_B > 0$, and so the free boundary is everywhere convex. As indicated in Fig. 6, the free boun-

[17]The theory of equations of this type, and extensive references, are given in Oleĭnik and Radkevič (1973). This book treats only the linear case [see Eq. (A18)], rather than the quasilinear case encountered in the leading logarithm model. An important difference found in the quasilinear case is that the location of the characteristic depends on the solution to the equation, rather than being *a priori* known. This is why the real characteristic of Eq. (2.72) behaves as a free, as opposed to a fixed, boundary.

dary intersects the axis of rotation at a point $\rho=0,z=z_A$; in Appendix A it is shown that $z_A>a$, with the possibility $z_A=a$ excluded. Apart from this statement, we have been unable to characterize analytically the structure of the free-boundary—rotation axis intersection.[18] Once Φ and $\mathbf{D}$ have been determined in the interior region, A^0 can be calculated from the formula

$$A^0(\rho,z)=-\int_0^z dz'\frac{\hat{z}\cdot\mathbf{D}(\rho,z')}{\varepsilon[D(\rho,z')]} . \tag{2.76}$$

An interesting alternative method, discussed in Appendix A, is to determine A^0 from the known solution for ε by solving the linear differential equation

$$\nabla\cdot(\varepsilon\nabla A^0)=-j^0 \tag{2.77}$$

within the free boundary.

Since Φ and $\mathbf{D}$ vanish identically outside the free boundary, continuity requires that ε remain zero in the whole of the exterior region. Thus the exterior scalar and vector potentials are constrained only by the requirement that the electric and magnetic fields satisfy the branch II condition

$$\mathbf{E}^2-\mathbf{B}^2=\kappa^2 , \tag{2.78}$$

but are otherwise undetermined. In other words, the functional L_{eff} has an infinite equivalence class of C^1 extrema $A^0,\mathbf{A}$, corresponding to all possible ways of satisfying Eq. (2.78) outside the free boundary. All members of this equivalence class[19] give the same A^0,Φ inside the free boundary, and make the same physical predictions.

The solution to the leading logarithm model is clearly qualitatively similar to the confinement domain found in the MIT "bag" model,[20] but there are important differences. At the boundary of an MIT "bag" the fields (the first derivatives of the potentials) are discontinuous, corresponding to the presence of a step function in the variational principle formulation. In the leading logarithm model, the boundary is a characteristic across which the fields are continuous, with only the first derivatives of the fields (the second derivatives of the potentials) having discontinuities. This behavior corresponds to the fact that the variational formulation of the leading logarithm model involves a smooth action functional L_{eff}.

[18]The numerical results given below suggest that the free boundary intersects the rotation axis at a right angle (rather than at a cusp), but we have no proof of this.

[19]A question which remains to be clarified in the field-theoretic context is whether the degenerate exterior solutions should be interpreted as vacuum structure. For articles advocating this view see, for example, Savvidy (1977), Pagels and Tomboulis (1978), and Nielsen (1981); for possible problems with this interpretation, see Kiskis (1980b) and Cabo and Shabad (1980).

[20]The MIT "bag" model was introduced by Chodos *et al.* (1974); for a review, see Hasenfratz and Kuti (1978), and for a reinterpretation within QCD, see Johnson (1978).

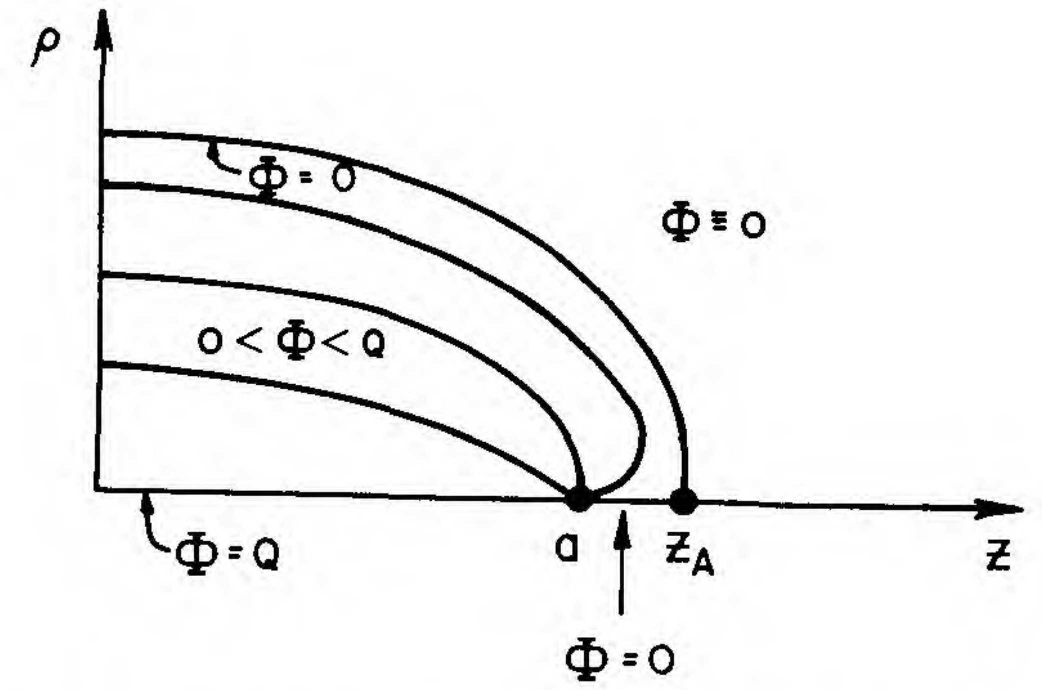


FIG. 6. Qualitative appearance of the solution of Eqs. (2.50), (2.56), and (2.59)–(2.61).

E. Axially symmetric Bogomol'nyi-Prasad-Sommerfield monopoles

As our final nonlinear example, let us consider a non-Abelian generalization of the Abelian Higgs model of Eq. (2.10), in which an SU(2) non-Abelian gauge field is coupled to a real scalar field φ^a, $a=1,2,3$, in the adjoint representation. The Lagrangian is

$$\begin{aligned} L&=\int d^3x\mathscr{L} , \\ \mathscr{L}&=\tfrac{1}{2}(\mathbf{E}^a\cdot\mathbf{E}^a-\mathbf{B}^a\cdot\mathbf{B}^a) \\ &\quad+\tfrac{1}{2}(\mathscr{D}_0\varphi^a)^2-\tfrac{1}{2}(\mathscr{D}_j\varphi^a)^2-\tfrac{1}{2}C[(\varphi^a)^2-\kappa^2]^2 , \end{aligned} \tag{2.79}$$

with the field strengths and covariant derivatives given by

$$\begin{aligned} E^{aj}&=-\mathscr{D}_jA^{a0}-\frac{\partial}{\partial t}A^{aj}, \\ B^{aj}&=\varepsilon^{jkl}\left[\frac{\partial}{\partial x^k}A^{al}+\tfrac{1}{2}\varepsilon^{abc}A^{bk}A^{cl}\right] , \end{aligned} \tag{2.80a}$$

$$\begin{aligned} \mathscr{D}_0\varphi^a&=\frac{\partial}{\partial t}\varphi^a-\varepsilon^{abc}A^{b0}\varphi^c , \\ \mathscr{D}_jw^a&=\frac{\partial}{\partial x^j}w^a+\varepsilon^{abc}A^{bj}w^c \text{ for } w^a=\varphi^a \text{ or } A^{a0} , \end{aligned} \tag{2.80b}$$

$$[\mathscr{D}_k,\mathscr{D}_j]w^a=\varepsilon^{kjl}\varepsilon^{abc}B^{bl}w^c . \tag{2.80c}$$

In writing Eqs. (2.79) and (2.80), we have for simplicity taken the gauge field coupling g to be unity. We will

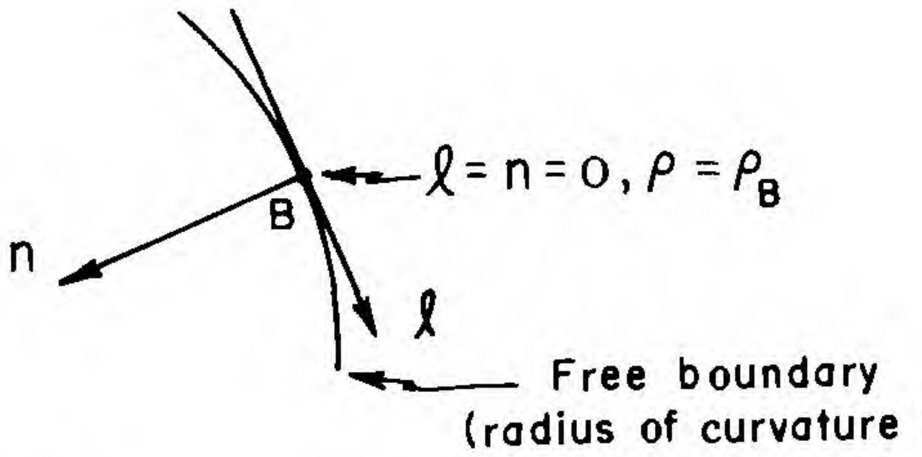


FIG. 7. Geometry near the free boundary used in Eqs. (2.75) and (A10)–(A12).

again be interested in static solutions for which all time derivatives vanish, but this time (since no external source charge-density coupling to A^{a0} has been included) we will take $A^{a0}=0$. After making these simplifications we have

$$-L=H=\int d^3x\mathscr{H}\,, \tag{2.81}$$

with $\mathscr{H}$ the field energy density

$$\mathscr{H}=\tfrac{1}{2}\mathbf{B}^a\cdot\mathbf{B}^a+\tfrac{1}{2}(\mathscr{D}_j\varphi^a)^2+\tfrac{1}{2}C[(\varphi^a)^2-\kappa^2]^2\,. \tag{2.82}$$

We will be interested in what follows in the finite energy extrema of Eq. (2.81), which clearly must satisfy

$$\lim_{r\to\infty}(\varphi^a)^2=\kappa^2\,. \tag{2.83}$$

Defining a unit SU(2) internal symmetry vector $\hat{n}^a(\hat{r})$ on the sphere at spatial infinity by

$$\begin{aligned}\hat{n}^a(\hat{r})&=\lim_{r\to\infty}\varphi^a(r\hat{r})/\kappa\,,\\ \hat{r}&=\mathbf{x}/r\,,\end{aligned} \tag{2.84}$$

we see that the complete specification of the boundary condition for φ^a requires specifying the number of times the two-sphere on which $\hat{n}^a$ lies is covered, when the two-sphere on which $\hat{r}$ lies is traversed once. This number (which must be an integer when $\hat{n}^a$ is a continuous function of $\hat{r}$) is the winding number or topological quantum number

$$n=\lim_{r\to\infty}\frac{1}{8\pi}\int_{\text{sphere at }\infty}d^2S^i\varepsilon^{ijk}\varepsilon^{abc}\hat{n}^a\frac{\partial}{\partial x^j}\hat{n}^b\frac{\partial}{\partial x^k}\hat{n}^c\,. \tag{2.85}$$

Hence the problem of finding finite energy extrema of H breaks up into discrete topological sectors.

Extrema of H for $C\neq 0$ are called 't Hooft (1974)–Polyakov (1974) monopoles. A very interesting special case, introduced by Prasad and Sommerfield (1975) and Bogomol'nyi (1976) and extensively studied since then,[21] is obtained by setting $C=0$ but retaining the boundary conditions of Eqs. (2.83)–(2.85), giving

$$\begin{aligned}&H=\int d^3x\tfrac{1}{2}(B^{aj}B^{aj}+\mathscr{D}_j\varphi^a\mathscr{D}_j\varphi^a)\,,\\ &\lim_{r\to\infty}\varphi^a=\kappa\hat{n}^a\,,\\ &\text{winding number of }\hat{n}^a=n\,.\end{aligned} \tag{2.86}$$

The Euler-Lagrange equations obtained by varying the functional H of Eq. (2.86) are

$$\begin{aligned}&\mathscr{D}_j\mathscr{D}_j\varphi^a=0\,,\\ &\varepsilon^{kjm}\mathscr{D}_jB^{am}=-\varepsilon^{abc}\varphi^b\mathscr{D}_k\varphi^c\,,\end{aligned} \tag{2.87}$$

while the field-potential relations of Eq.(2.80) imply that

$$\begin{aligned}&\mathscr{D}_jB^{aj}=0\,,\\ &\varepsilon^{kjm}\mathscr{D}_j\mathscr{D}_m\varphi^a=-\varepsilon^{abc}\varphi^bB^{ck}\,.\end{aligned} \tag{2.88}$$

Although Eqs. (2.87) are complicated second-order differential equations, they are satisfied by any φ^a, A^{aj} satisfying the first-order differential equations

$$-\mathscr{D}_j\varphi^a=\xi B^{aj}\,,\quad \xi=\pm 1\,, \tag{2.89}$$

with the cases $\xi=1\ (-1)$ termed, respectively, self-dual (anti-self-dual).[22] To see that Eq. (2.89) suffices, we note that by Eq. (2.88) we have

$$\mathscr{D}_j\mathscr{D}_j\varphi^a=-\xi\mathscr{D}_jB^{aj}=0\,, \tag{2.90a}$$

while by using Eq. (2.80c) we get

$$\begin{aligned}\varepsilon^{kjm}\mathscr{D}_jB^{am}&=-\xi\varepsilon^{kjm}\mathscr{D}_j\mathscr{D}_m\varphi^a\\ &=\xi\varepsilon^{abc}\varphi^bB^{ck}=-\xi^2\varepsilon^{abc}\varphi^b\mathscr{D}_k\varphi^c\,.\end{aligned} \tag{2.90b}$$

Remarkably, Eq. (2.89) is also a necessary condition for a minimum of H, as may be seen by the following[23] rearrangement of Eq. (2.86),

$$\begin{aligned}H&=H_1+H_2\,,\\ H_1&=\int d^3x\tfrac{1}{2}(B^{aj}+\xi\mathscr{D}_j\varphi^a)^2\,,\\ H_2&=-\xi\int d^3xB^{aj}\mathscr{D}_j\varphi^a\\ &=-\xi\int d^3x\mathscr{D}_j(B^{aj}\varphi^a)\\ &=-\xi\int_{\text{sphere at }\infty}d^2S^jB^{aj}\varphi^a\,.\end{aligned} \tag{2.91}$$

Since H_2 reduces to a surface term, the functional H can be extremal only for fields for which H_1 vanishes, giving the condition of Eq. (2.89). The surface term H_2 can be evaluated by noting that in order for the integral of Eq. (2.86) to converge, $\mathscr{D}_j\hat{n}^a$ must vanish at large r, giving the following relation between A^{aj} and $\hat{n}^a$ on the sphere at infinity,

$$A^{aj}=\hat{n}^a\hat{n}^bA^{bj}-\varepsilon^{abc}\hat{n}^b\frac{\partial}{\partial x^j}\hat{n}^c\,. \tag{2.92}$$

Combining Eqs. (2.80a) and (2.92), one finds, after some algebra,

$$B^{aj}\hat{n}^a=\frac{\partial}{\partial x^k}(\varepsilon^{jkl}\hat{n}^aA^{al})-\tfrac{1}{2}\varepsilon^{jkl}\varepsilon^{abc}\hat{n}^a\frac{\partial}{\partial x^k}\hat{n}^b\frac{\partial}{\partial x^l}\hat{n}^c\,. \tag{2.93}$$

---

[21]For recent reviews on monopoles, see Jaffe and Taubes (1980) and O'Raifeartaigh and Rouhani (1981).

[22]This terminology stems from the fact that $-\mathscr{D}_j\varphi^a$ can be formally reinterpreted as a static Euclidean electric field strength $E^{aj}_{(E)}\equiv-\mathscr{D}_jA^{a0}_{(E)}$, $A^{a0}_{(E)}\equiv\varphi^a$, in terms of which $\mathscr{L}=-\tfrac{1}{2}(\mathbf{B}^a\cdot\mathbf{B}^a+\mathbf{E}^a_{(E)}\cdot\mathbf{E}^a_{(E)})$. The self-dual ($\xi=1$) solutions satisfy $E^{aj}_{(E)}=B^{aj}$, while the anti-self-dual ($\xi=-1$) solutions satisfy $E^{aj}_{(E)}=-B^{aj}$. Clearly, a $\xi=1$ solution can be converted to a $\xi=-1$ solution simply by changing the sign of φ^a.

[23]This argument is due to Bogomol'nyi (1976) and to Coleman, Parke, Neveu, and Sommerfield (1977).

The surface integral of the first term on the right-hand side of Eq. (2.93) vanishes; and so, referring back to Eq. (2.85), we get

$$H_2=4\pi\kappa\xi n \ . \tag{2.94}$$

Since the positivity of H implies that H_2 is positive when H_1 vanishes, we conclude that

$$H_2=4\pi\kappa|n| \ , \ \ \xi=n/|n| \ . \tag{2.95}$$

Hence the minimum of H in each topological sector is determined solely by the topological quantum number.

The minimum of H in the $n=1$ topological sector is given by the simple expression

$$\begin{aligned} \varphi^a &= -\frac{x^a}{r^2}(1-\kappa r\coth\kappa r) \ , \\ A^{aj} &= \frac{\varepsilon^{ajl}x^l}{r^2}\left[1-\frac{\kappa r}{\sinh\kappa r}\right] , \end{aligned} \tag{2.96}$$

which satisfies Eqs. (2.89) with $\xi=1$. This solution is the simplest of a family of solutions of Eq. (2.89), in which the Higgs field φ^a and the potentials A^{aj} are axially symmetric and reflection symmetric, as described by the following ansatz:

$$\begin{aligned} \varphi^a &= h_1\hat{z}^a+h_2\hat{\rho}_n^a \ , \\ A^{aj} &= \hat{\phi}^j\left[\frac{f_1-n}{\rho}\hat{z}^a+\frac{f_2}{\rho}\hat{\rho}_n^a\right]-(\hat{z}^j a_1+\hat{\rho}^j a_2)\hat{\phi}_n^a \ , \\ \rho &= (x^2+y^2)^{1/2}, \ \ \phi=\tan^{-1}(y/x) \ , \\ \hat{z} &= (0,0,1) \ , \\ \hat{\rho}_n &= (\cos n\phi,\sin n\phi,0), \ \ \hat{\rho}=\hat{\rho}_1 \ , \\ \hat{\phi}_n &= (-\sin n\phi,\cos n\phi,0), \ \ \hat{\phi}=\hat{\phi}_1 \ . \end{aligned} \tag{2.97}$$

The potentials $h_{1,2}$, $f_{1,2}$, and $a_{1,2}$ are functions of ρ and z, with $z\leftrightarrow -z$ reflection symmetry and behavior on the $z=0$ and $\rho=0$ axes as follows:

$$\begin{aligned} &h_2,f_1,a_1 \text{ even in } z \ , \\ &h_1,f_2,a_2 \text{ odd in } z \ , \\ &h_1=f_2=a_2=0 \text{ at } z=0 \ , \\ &a_1=f_2=h_2=0, \ f_1=n \text{ at } \rho=0, \ n\geq 1 \ . \end{aligned} \tag{2.98}$$

Although analytic forms for the solutions of Eq. (2.89) within the ansatz of Eqs. (2.97) and (2.98) are now known,[24] we will treat the problem of finding the axially symmetric, reflection-symmetric minima of H as a numerical example in Sec. IV.D. In Appendix B we give explicit expressions for the functional H when expressed in terms of the potentials $h_{1,2},\ldots$, and discuss the residual Abelian gauge invariance of the six-function ansatz and the choice of a gauge-fixing term for numerical work. From the equations given in Appendix B, a straightforward calculation shows that the leading behavior of the potentials at infinity[25] is given by the following formulas (in which terms of order $e^{-\kappa r}$ are omitted):

$$\begin{aligned} &h_1=h\cos\vartheta, \ h_2=h\sin\vartheta \ , \\ &a_1=-\frac{\sin\vartheta}{r}, \ a_2=\frac{\cos\vartheta}{r} \ , \\ &f_1=f\cos\vartheta, \ f_2=f\sin\vartheta \ , \\ &h=\kappa-\frac{n}{r}+\sum_{\substack{l=2\\ l \text{ even}}}^{\infty}\frac{a_l^{(\infty)}P_l(\cos\vartheta)}{r^{l+1}} \ , \\ &f=n\cos\vartheta+\rho\sum_{\substack{l=2\\ l \text{ even}}}^{\infty}\frac{a_l^{(\infty)}P_{l1}(\cos\vartheta)}{lr^{l+1}} \ , \\ &r=(\rho^2+z^2)^{1/2}, \ \vartheta=\tan^{-1}(\rho/z) \ . \end{aligned} \tag{2.99}$$

The leading behavior at the origin[25] is also calculated in Appendix B, where it is shown that even in the higher topological sectors $n>1$, the Higgs field φ^a has only a first-order zero at $r=0$.

III. RELAXATION METHODS FOR THE NUMERICAL SOLUTION OF ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

A. Introduction

In this section we give a detailed introduction to both the theoretical and practical aspects of the numerical solution of elliptic partial differential equations of the type encountered in Sec. II. We assume that the reader has read Sec. II.A and especially Sec. II.B, and has at least glanced over Secs. II.C–II.E. However, the discussion of numerical methods which follows is essentially self-contained, and makes only minimal references back to the formulas in Sec. II. Motivated by the fact that the numerical solution of the models of Secs. II.C–II.E can be reduced to a sequence of solutions to linear differential

[24]For the analytic construction of axially symmetric multimonopole solutions and references, see Prasad and Rossi (1981).

[25]Equations (2.99) and (B11) are the expansions calculated in the gauge $\partial_z a_1+\partial_\rho a_2=0$. For $n=1$, this gauge condition does not uniquely fix the order r terms in $a_{1,2}$ near $r=0$; the general solution for $n=1$ is $a_1=-b\rho(\frac{1}{2}+\alpha)+O(r^2)$, $a_2=bz(\frac{1}{2}-\alpha)+O(r^2)$, with α a free parameter. The $n=1$ case of Eq. (B11) corresponds to $\alpha=\frac{1}{2}$, while the standard form of the single monopole solution given in Eq. (2.96) corresponds to $\alpha=0$. [To put Eq. (2.96) in the form of Eq. (2.97), one uses $\hat{r}^a=\hat{z}^a\cos\vartheta+\hat{\rho}^a\sin\vartheta$, $\varepsilon^{ajl}\hat{r}^l=(\hat{\rho}^a\hat{\varphi}^j-\hat{\rho}^j\hat{\varphi}^a)\cos\vartheta+(\hat{\varphi}^a\hat{z}^j-\hat{\varphi}^j\hat{z}^a)\times\sin\vartheta$.] The expansions of Eqs. (2.99) and (B11) were derived by Adler (unpublished), Rebbi and Rossi (1980), and Houston and O'Raifeartaigh (1981).

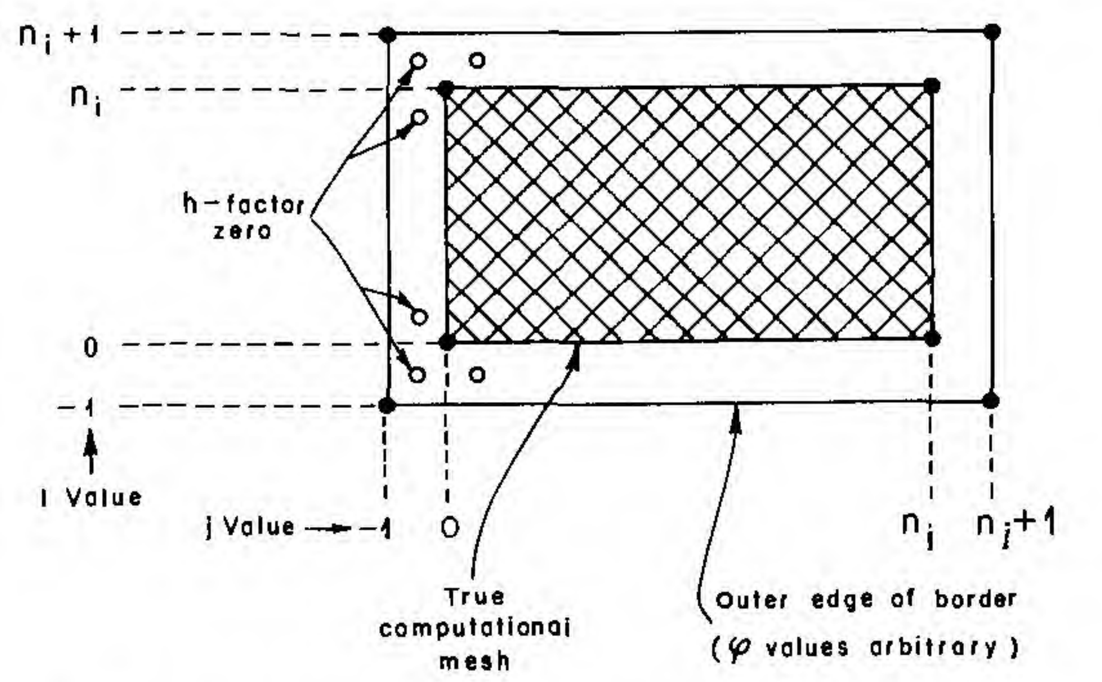


FIG. 10. Computational mesh extended by one unit cell to form a border. At the half nodes within the border, the h factors of Table I are assigned the value zero. The corresponding summation limits for the four coverings of Fig. 9 are given in Table II.

signed to φ at nodes on the outer edge of the border, since these nodes automatically appear multiplied by an h factor of zero. When this procedure is used, the summation limits for the four coverings of Figs. 9(a)–9(d) are as given in Table II.

C. Iterative methods of solution

Let us now suppose that the discretization procedure of the preceding section has been carried out on a computational lattice with n_i+1, n_j+1 nodes in the ρ, z directions, respectively. Equation (3.25) then gives us a set of $N=(n_i+1)(n_j+1)$ linear equations in the N unknown variables $\varphi_{i,j}$. Since N is typically a very large number (up to $\approx 4\times 10^4$ in the computations described in Sec. IV), the direct solution of this set of equations by matrix inversion is not feasible, and we must resort instead to an iterative method of solution.

Before describing the specific algorithms used to solve Eq. (3.25), let us first discuss a simple and familiar iterative method which will also be needed in the applications of Sec. IV. This is the Newton iteration for finding the roots of the equation

$$f(w)=0\,, \tag{3.33}$$

which is constructed as follows. Given an estimate $w^{(n)}$ of a root w^* of Eq. (3.33), we Taylor expand $f(w)$ around $w^{(n)}$, giving

$$f(w)=f(w^{(n)})+(w-w^{(n)})f'(w^{(n)}) + \tfrac{1}{2}(w-w^{(n)})^2 f''(w^{(n)})+\cdots\,. \tag{3.34}$$

When substituted into Eq. (3.33), Eq. (3.34) gives an exact power-series equation for w^*. When Eq. (3.34) is approximated by the first two terms in the series expansion, it gives a new approximation to w^*,

$$w^*\approx w^{(n+1)}=w^{(n)}-\frac{f(w^{(n)})}{f'(w^{(n)})}\,. \tag{3.35}$$

The error of the new approximation can be estimated by subtracting the two equations

$$0=f(w^{(n)})+(w^{(n+1)}-w^{(n)})f'(w^{(n)})\,, \tag{3.36a}$$

$$0=f(w^{(n)})+(w^*-w^{(n)})f'(w^{(n)})+\tfrac{1}{2}(w^*-w^{(n)})^2 f''(w^{(n)}) +O[(w^*-w^{(n)})^3]\,,$$

giving

$$\begin{aligned} w^{(n+1)}-w^* &\approx \tfrac{1}{2}(w^{(n)}-w^*)^2\frac{f''(w^{(n)})}{f'(w^{(n)})}+O[(w^{(n)}-w^*)^3] \\ &\approx \tfrac{1}{2}(w^{(n)}-w^*)^2\frac{f''(w^*)}{f'(w^*)}+O[(w^{(n)}-w^*)^3]\,. \end{aligned} \tag{3.36b}$$

Hence the error after $n+1$ iterations is of order the square of the error after n iterations, and convergence proceeds very rapidly to w^*, provided that the initial guess $w^{(0)}$ lies close enough to w^*. Rewriting Eq. (3.36b) as

$$\frac{|w^{(1)}-w^*|}{|w^{(0)}-w^*|}\approx\tfrac{1}{2}\,|w^{(0)}-w^*|\,\left|\frac{f''(w^*)}{f'(w^*)}\right| +O[(w^{(0)}-w^*)^2]\,, \tag{3.37a}$$

we see that a sufficient condition on $w^{(0)}$ to guarantee that the Newton iteration converges to the root w^* is

TABLE II. Summation limits for the coverings of Fig. 9, using the "bordered" mesh of Fig. 10.

Unit cell	Lower i,j limits	Upper i limit	Upper j limit
"ρ-derivative covering" Fig. 9(a)	0	n_i-1	n_j
"z-derivative covering" Fig. 9(b)	0	n_i	n_j-1
"nonderivative covering" Fig. 9(c)	0	n_i	n_j
"ρ,z derivative covering" Fig. 9(d)	0	n_i-1	n_j-1

$$\tfrac{1}{2}\,|\,w^{(0)}-w^*\,|\,\left|\frac{f''(w^*)}{f'(w^*)}\right| \ll 1\,. \tag{3.37b}$$

If Eq. (3.33) has several roots $w_1^*, w_2^*, \ldots$, there will be an interval around each root w_i^* within which the Newton iteration converges to that root.

Let us now proceed to the simplest case in which we encounter the iterative method which will be used to solve Eq. (3.25). We consider the discrete functional

$$L=\tfrac{1}{2}\sum_{r=1}^{N}\sum_{s=1}^{N}A_{rs}\varphi_r\varphi_s-\sum_{r=1}^{N}J_r\varphi_r\,, \tag{3.38}$$

with A_{rs} a real, symmetric matrix with positive eigenvalues. Since the matrix A_{rs} is invertible, L attains a unique minimum when the nodal variables $\varphi_r, r=1,\ldots,N$ satisfy the N linear equations

$$0=\frac{\partial L}{\partial \varphi_r} \Longrightarrow \sum_{s=1}^{N}A_{rs}\varphi_s=J_r\,. \tag{3.39}$$

An iterative method for finding this minimum can now be constructed as follows. Let us repeatedly sweep through the φ_r, proceeding from φ_1 to φ_N and then starting over again with φ_1,

$$\underbrace{\varphi_1,\ldots,\varphi_N}_{\text{first sweep}},\underbrace{\varphi_1,\ldots,\varphi_N}_{\text{second sweep}},\ldots\,, \tag{3.40}$$

at each step replacing the variable φ_R being considered by the value which minimizes L when all other variables $\varphi_r, r\neq R$ are held fixed. Specifically, let $\varphi_r^{(n)}$ be the values of all the nodal variables when the sweep reaches the variable φ_R, so that at this stage L has the value

$$L^{(n)}=\tfrac{1}{2}\sum_{r=1}^{N}\sum_{s=1}^{N}A_{rs}\varphi_r^{(n)}\varphi_s^{(n)}-\sum_{r=1}^{N}J_r\varphi_r^{(n)}\,. \tag{3.41}$$

Regarding L as a function of the single variable φ_R, with the other variables fixed, we have

$$L=\tfrac{1}{2}\varphi_R^2A_{RR}+\varphi_R\left[\sum_{s\neq R}A_{Rs}\varphi_s^{(n)}-J_R\right]+\tfrac{1}{2}\sum_{r\neq R}\sum_{s\neq R}A_{rs}\varphi_r^{(n)}\varphi_s^{(n)}-\sum_{r\neq R}J_r\varphi_r^{(n)}\,. \tag{3.42}$$

Choosing $\varphi_R^{(n+1)}$ to minimize Eq. (3.42) with respect to φ_R, we get

$$\varphi_R^{(n+1)}=-\frac{1}{A_{RR}}\left[\sum_{s\neq R}A_{Rs}\varphi_s^{(n)}-J_R\right]=\varphi_R^{(n)}+\Delta\varphi_R^{(n)}\,,$$
$$\Delta\varphi_R^{(n)}=-\frac{1}{A_{RR}}\left[\sum_{s}A_{Rs}\varphi_s^{(n)}-J_R\right], \tag{3.43}$$

giving as the corresponding change in the action

$$\begin{aligned}L^{(n+1)}-L^{(n)}&=\tfrac{1}{2}[(\varphi_R^{(n+1)})^2-(\varphi_R^{(n)})^2]A_{RR}+(\varphi_R^{(n+1)}-\varphi_R^{(n)})\left[\sum_{s\neq R}A_{Rs}\varphi_s^{(n)}-J_R\right]\\&=\tfrac{1}{2}\Delta\varphi_R^{(n)}(\Delta\varphi_R^{(n)}+2\varphi_R^{(n)})A_{RR}+\Delta\varphi_R^{(n)}(-A_{RR}\Delta\varphi_R^{(n)}-A_{RR}\varphi_R^{(n)})\\&=-\tfrac{1}{2}A_{RR}(\Delta\varphi_R^{(n)})^2<0\,.\end{aligned} \tag{3.44}$$

(In the final line we have used the fact that since the matrix A_{rs} has positive eigenvalues, the diagonal matrix elements A_{RR} are all positive.) Hence, under the iteration of Eq. (3.40), the Lagrangian is monotone decreasing. In problems of physical interest with positive definite A_{rs}, we expect the Lagrangian with the source term included to be bounded from below. Equations (3.44) and (3.43) then guarantee that in the limit as n becomes infinite, the $L^{(n)}$'s converge to the minimum value of L, while the $\varphi_r^{(n)}$'s converge to the solution of Eq. (3.39). This method of solving the system of Eq. (3.39) is known as the Gauss-Seidel iteration.

As we have just seen, in the Gauss-Seidel iteration each nodal variable is successively relaxed to the value which, at that stage of the iteration, minimizes L. An important variant of the basic method, called the successively over-relaxed (SOR) Gauss-Seidel iteration, is defined by the recipe

$$\begin{aligned}\varphi_R^{(n)}\to\varphi_R^{(n+1),\mathrm{SOR}}&=\varphi_R^{(n)}+\omega\Delta\varphi_R^{(n)}\\&=\omega\varphi_R^{(n+1)}+(1-\omega)\varphi_R^{(n)},\quad \omega\geq 1\,,\end{aligned} \tag{3.45}$$

with $\varphi_R^{(n+1)}$ and $\Delta\varphi_R^{(n)}$ given by Eq. (3.43). In other words, instead of relaxing $\varphi_R^{(n)}$ to the value which minimizes L, one systematically overshoots beyond this value. Using the over-relaxed iteration, the change in the Lagrangian is

$$\begin{aligned}L^{(n+1),\mathrm{SOR}}-L^{(n)}&=\tfrac{1}{2}[(\varphi_R^{(n+1),\mathrm{SOR}})^2-(\varphi_R^{(n)})^2]A_{RR}+(\varphi_R^{(n+1),\mathrm{SOR}}-\varphi_R^{(n)})\left[\sum_{s\neq R}A_{Rs}\varphi_s^{(n)}-J_R\right]\\&=\tfrac{1}{2}\omega\Delta\varphi_R^{(n)}(\omega\Delta\varphi_R^{(n)}+2\varphi_R^{(n)})A_{RR}+\omega\Delta\varphi_R^{(n)}(-A_{RR}\Delta\varphi_R^{(n)}-A_{RR}\varphi_R^{(n)})\\&=-\tfrac{1}{2}\omega(2-\omega)A_{RR}(\Delta\varphi_R^{(n)})^2\,.\end{aligned} \tag{3.46}$$

Thus provided that

$$0<\omega<2\ , \tag{3.47}$$

the Lagrangian remains monotone decreasing, and the SOR iteration still converges to the solution of Eq. (3.39).

For a general, symmetric matrix A_{rs}, each individual iteration step in Eq. (3.43) involves evaluating a sum of N terms, which would be computationally costly for large N. However, in the specific problems to which we will apply iterative methods, r and s are composite indices

$$\begin{aligned} r&=(i,j)\ ,\\ s&=(i',j')\ , \end{aligned} \tag{3.48}$$

denoting nodes of the computational lattice, and A_{rs} is a sparse matrix in which A_{Rs} is nonvanishing only for a small number of index values s for each R. Because we have followed the prescription of squaring before averaging to ensure that the discretized action contains only nearest-neighbor couplings, we find in fact that A_{Rs} is nonvanishing only for the five index values s corresponding to the node R and its four nearest neighbors. Hence each iteration step involves only a short computation which is independent of the size of the computational lattice. If the sweep is performed[28] in "typewriter ordering"

$$\begin{aligned} &(i,j){:}(0,0),(1,0),\ldots,(n_i,0),(0,1),(1,1),\ldots,(n_i,1)\ ,\\ &\ldots,(0,n_j),(1,n_j),\ldots,(n_i,n_j)\ , \end{aligned} \tag{3.49}$$

then the over-relaxed iteration for Eq. (3.25) is

$$\begin{aligned} \varphi_{i,j}^{(n)}\to\varphi_{i,j}^{(n+1),\mathrm{SOR}}&=\omega\varphi_{i,j}^{(n+1)}+(1-\omega)\varphi_{i,j}^{(n)}\ ,\\ \varphi_{i,j}^{(n+1)}&=\left[\left[\frac{\Delta z}{\Delta\rho}+\frac{\Delta\rho}{\Delta z}\right](h_{i+1/2,j+1/2}+h_{i+1/2,j-1/2}+h_{i-1/2,j+1/2}+h_{i-1/2,j-1/2})\right]^{-1}\\ &\quad\times\left\{\frac{\Delta z}{\Delta\rho}(h_{i+1/2,j+1/2}+h_{i+1/2,j-1/2})\varphi_{i+1,j}^{(n)}+\frac{\Delta z}{\Delta\rho}(h_{i-1/2,j+1/2}+h_{i-1/2,j-1/2})\varphi_{i-1,j}^{(n+1)}\right.\\ &\quad+\frac{\Delta\rho}{\Delta z}(h_{i+1/2,j+1/2}+h_{i-1/2,j+1/2})\varphi_{i,j+1}^{(n)}+\frac{\Delta\rho}{\Delta z}(h_{i+1/2,j-1/2}+h_{i-1/2,j-1/2})\varphi_{i,j-1}^{(n+1)}\\ &\quad\left.+\frac{Q}{\pi}\delta_{i,0}(\delta_{j,n_Q}-\delta_{j,-n_Q})\right\},\quad h_{i+1/2,j+1/2}\equiv\rho_{i+1/2}\epsilon_{i+1/2,j+1/2}\ . \end{aligned} \tag{3.50}$$

Equation (3.50) applies to all i,j in the range $0\le i\le n_i\ ,0\le j\le n_j$, since, by virtue of the "bordering" procedure discussed above in Sec. III.B, all nodal values with indices outside this range appear in Eq. (3.50) multiplied by vanishing h factors. An initial guess $\varphi_{i,j}^{(0)}$ must be supplied as an input to the iterative process. In practice, to achieve a poor-man's version of the "hierarchical" iterative schemes (see below), we follow the procedure of first iterating to convergence on a very coarse mesh starting from a specified $\varphi_{i,j}^{(0)}$, which is chosen to be reasonably close (without sacrificing simplicity)[29] to the anticipated solution of the problem. We then successively double the mesh and iterate to convergence, taking as the new initial guess after each doubling a linear interpolation of the converged $\varphi_{i,j}$ values on the preceding coarser mesh.

According to the discussion of Eqs. (3.41)–(3.47), an iteration with $\omega=1$ produces the largest possible *single-step* reduction in the Lagrangian L. Hence at the beginning of an iterative solution one always makes three to ten complete passes through the computational lattice with $\omega=1$ (or with an ω which is gradually increased starting from 1) to eliminate the largest deviations between the initial guess $\varphi_{i,j}^{(0)}$ and the fully converged solution $\varphi_{i,j}^{(\infty)}$. After these initial iterations, the optimal strategy is to use an ω value larger than 1. The reason is that a general analysis[30] of the iterative process shows that in the asymptotic large-n limit, the difference $\varphi_{i,j}^{(n)}-\varphi_{i,j}^{(\infty)}$ behaves as

$$\varphi_{i,j}^{(n)}-\varphi_{i,j}^{(\infty)}\sim ce^{-n\gamma(\omega)}\ , \tag{3.51}$$

with the decay constant $\gamma(\omega)$ attaining its maximum at an ω value $\omega=\omega_{\mathrm{opt}}\ ,1<\omega_{\mathrm{opt}}<2$. Hence one clearly achieves the maximum rate of convergence by letting ω tend to ω_{opt} from below after the initial iterations. Values of ω larger than ω_{opt} should be avoided, since in general they produce slower convergence than values of ω an equivalent distance below ω_{opt}, and since they can lead to instabilities in nonlinear problems. The optimum value of ω can be estimated from the formula [Garabedian (1956)]

$$\omega_{\mathrm{opt}}\approx\frac{2}{1+Ch}\ , \tag{3.52}$$

[28]This terminology has been borrowed from Hockney and Eastwood (1981), who discuss alternative sweeps as well.

[29]In practice, there is no great gain in convergence to be achieved by using elaborate functional forms in the initial guess.

[30]See the books cited at the end of Sec. III.A for details.

with

$$h=(n_i n_j)^{-1/2} \tag{3.53}$$

a measure of the fineness of the computational lattice, and with C a constant which depends on the lattice geometry and the boundary conditions. (For the two-dimensional Laplace equation in rectangular coordinates with Dirichlet boundary conditions, $C \sim 3$.) An empirical method for estimating the value of $\omega_{\rm opt}$ is discussed below in Sec. III.G; from the general form of Eq. (3.52) we infer the useful fact that when the computational mesh is doubled, so that

$$\begin{aligned} &n_i \rightarrow 2n_i, \quad n_j \rightarrow 2n_j \ , \\ &h \rightarrow h/2 \ , \end{aligned} \tag{3.54}$$

the corresponding change in $\omega_{\rm opt}$ is

$$\omega_{\rm opt} \rightarrow \frac{4\omega_{\rm opt}}{2+\omega_{\rm opt}} \ . \tag{3.55}$$

An intuitive (and mathematically correct) way to visualize the over-relaxation algorithm is to think of n as a time step and of the algorithm as a time-dependent dissipative process, with a steady-state equilibrium at the converged solution $\varphi_{i,j}^{(\infty)}$. The starting guess $\varphi_{i,j}^{(0)}$ in general deviates from $\varphi_{i,j}^{(\infty)}$ by both large localized transients and by relatively smooth errors. The initial iterations with $\omega=1$ are used to rapidly eliminate the localized transients. The later iterations with $\omega=\omega_{\rm opt}$ minimize the relatively long time constant $[\gamma(\omega)]^{-1}$ with which the smooth errors damp away. The optimal use of the over-relaxation method requires attention to eliminating both localized transients and smooth (or long-range) errors. One method of doing this in a systematic way is the "Chebyshev acceleration" method described by Hockney and Eastwood (1981), in which the lattice is scanned using an "even-odd checkerboard ordering" (as opposed to "typewriter ordering") and in which ω is incremented from $\omega=1$ to $\omega=\omega_{\rm opt}$ in a prescribed way at the beginning of each even-semilattice and each odd-semilattice sweep. A second way of accomplishing this is through various "hierarchical" schemes[31] in which the full computational lattice is scanned in only a fraction of the sweeps, with the remaining sweeps used to scan sublattices of the basic lattice, constructed by using a larger unit cell containing two, four, etc., fundamental unit cells. The Chebyshev acceleration and hierarchical schemes can be proved to be optimal ones, according to well-defined criteria, for solving elliptical partial differential equation problems based on the Laplacian (∇^2) and similar linear operators. Since for nonlinear problems it usually is not possible explicitly to construct an optimal algorithm, we use instead a simpler method which is effectively equivalent. As discussed above, what we do is to start the iteration on a very small (typically 7×7) computational lattice, iterate to convergence, and then use an interpolation of this solution as the initial guess for iteration on a computational mesh which has been doubled as in Eq. (3.54), and so forth. After each doubling, ω is reset to 1 for several iterations to eliminate transients arising from the interpolation (these are strongly evident in the unit cells along the axis of rotation) and then increased to the $\omega_{\rm opt}$ appropriate to the new mesh spacing h. This procedure gives very satisfactory convergence and automatically generates a sequence of fully converged L values on progressively finer meshes, permitting an examination of the convergence of the discrete solution as the mesh spacing h approaches zero.

Let us consider next the imposition of boundary conditions in carrying out the iterative solution. From Eqs. (3.1)–(3.3), we see that the differential equation and boundary conditions of our dielectric model are invariant, and the source current j^0 changes sign, under the reflection operation $z \rightarrow -z$. This implies that the solution φ also has odd reflection symmetry, and vanishes on the equatorial plane $z=0$. Although this symmetry emerges automatically if the boundary value problem of Eqs. (3.1)–(3.3) is solved over the full physical space $0 \le \rho < \infty, -\infty < z < \infty$, we can clearly save computer time if we impose the symmetry at the outset, by solving instead a boundary value problem on the half space $0 \le \rho < \infty, 0 \le z < \infty$. Similarly, if the source current j^0 were replaced by

$$j'^0 = Q\delta(x)\delta(y)[\delta(z-a)+\delta(z+a)] \ , \tag{3.56}$$

which is invariant under the reflection $z \rightarrow -z$, then the corresponding solution φ would have even reflection symmetry, and $\partial_z\varphi$ would vanish at $z=0$. Again, although this symmetry would emerge automatically from the full-space boundary-value problem, we can halve the computational effort by using the symmetry to reduce the computation to an equivalent half-space boundary-value problem.

When we solve the numerical problem on a half space, we introduce an inner boundary $z=0, 0 \le \rho < \infty$ on which a boundary condition must be specified, together with the outer boundary condition of Eq. (3.3). On this inner boundary, the appropriate boundary conditions are, respectively, the Dirichlet or Neumann conditions,

$$\varphi=0 \text{ at } z=0,\ j^0 \text{ odd [Eq. (3.1)]} \ , \tag{3.57a}$$

$$\partial_z\varphi=0 \text{ at } z=0,\ j^0 \text{ even [Eq. (3.56)]} \ , \tag{3.57b}$$

and we must translate each of these into a corresponding updating algorithm for the lattice nodes on the line $z=0$. (The more general Robin or mixed boundary condition $\alpha\varphi+\beta\partial_z\varphi=0$ can also be implemented computationally, but will not be encountered in any of the models studied in this paper.) In addition, in either the full-space or half-space problems, there is an inner coordinate boundary

---

[31]Hierarchical over-relaxation methods have been discussed, for example, by Brandt (1977) and Press (1978).

the iteration will depend strongly on what is being measured at the end of the iterative process.

IV. NUMERICAL SOLUTION OF THE MODELS OF SEC. II

A. Introduction

Let us proceed now to apply the numerical methods described in Sec. III to the nonlinear models formulated in Sec. II. In Table III we summarize which dependent variables in the three models are discretized on the node lattice, and which are discretized on the half-node lattice, together with the boundary conditions which are imposed during iteration. In the brief sections which follow we discuss aspects of the numerical analysis which are specific to the three models and give sample numerical results.

B. The Abelian Higgs model

Following the analysis of Sec. III.E and Appendix C, we explicitly subtract off the Coulomb self-energy from L by making the substitution

$$A^0 = A_C^0 + B^0 , \tag{4.1}$$

with A_C^0 the Coulomb potential of Eq. (C1) and with B^0 a new dependent variable. Because A_C^0 and φ are both singular at the charges [cf. Eq. (2.17)], the charge coordinate $z=a$ is taken to lie midway between nodes of the computational lattice:

$$a=(n_Q+\tfrac{1}{2})\Delta z , \tag{4.2}$$

with n_Q an integer. We choose the unit of length so that $\kappa=1$, giving $\varphi\rightarrow 1$ as the boundary condition on φ at infinity. Since this boundary condition follows from requiring L to be extremal (L is infinite if $\varphi \not\rightarrow 1$ at infinity), it can be enforced computationally by simply iterating the nodal values for φ which lie on the outer boundary of the computational mesh. An alternative procedure would be to set $\varphi=1$ on the outer boundary; the two methods give the same result in the limit as $\rho_{\max}\rightarrow\infty, z_{\max}\rightarrow\infty$, but the iterative boundary condition is preferable for finite meshes.[36] To get a good approximation to the infinite volume solution, $\rho_{\max}$ and $z_{\max}$ must be chosen large compared with the characteristic exponential decay lengths appearing in Eq. (2.17), requiring (for $\kappa=1$) that

$$\min(\rho_{\max},z_{\max}) \gg \max[(2C)^{-1/2},(2e^2)^{-1/2}] . \tag{4.3}$$

Sample results for the Abelian Higgs model, calculated with

$$\kappa=C=e=Q^2/(4\pi)=1 , \qquad \rho_{\max}=z_{\max}=3, \quad a=1.625 , \tag{4.4}$$

are shown in Figs. 12(a)–12(d). These figures give values of φ and A^0 (plotted vertically) on a plane passing through the axis of rotation (represented by the horizontal plane in the figures). One can see clearly the peaks in φ and A^0 at the charges, as well as the exponential decay of φ towards 1 and of A^0 towards 0 at infinity. In Figs. 12(c) and 12(d), in which the vertical scale has been magnified by a factor of 10, one can also see that the structure of the solution extends to the computational boundary, in marked contrast to the behavior found below in the solution of the leading logarithm model.

C. The leading logarithm model

In discretizing the leading logarithm model, we put the charge coordinate $z=a$ on a node of the computational lattice,

$$a=n_Q\Delta z , \tag{4.5}$$

with n_Q an integer, and enforce the step function boundary condition of Eq. (2.50) by requiring

$$\begin{aligned} &\Phi_{0,j}=Q, \quad 0\le j<n_Q , \\ &\Phi_{0,j}=\tfrac{1}{2}Q, \quad j=n_Q , \\ &\Phi_{0,j}=0, \quad n_Q<j\le n_j . \end{aligned} \tag{4.6}$$

[An alternative procedure would be to put the charge coordinate midway between lattice nodes by taking $a=(n_Q+\frac{1}{2})\Delta z$, giving the boundary condition $\Phi_{0,j}=Q, 0\le j\le n_Q$ and $\Phi_{0,j}=0, n_Q+1\le j\le n_j$.] Because the solution for Φ is confined within a finite free boundary, the numerical solution is independent of $\rho_{\max},z_{\max}$, provided that these are large enough for the fully converged free boundary to lie entirely within the computational mesh. To facilitate picking values of $\rho_{\max},z_{\max}$ which are large enough to contain the free boundary but are not excessively so, we have included a control parameter option in the program which permits the adjustment of the limits of the computational mesh during iteration.

In carrying out the iteration we do not let the dielectric function ε assume the value 0, but rather impose a minimum value $\varepsilon_{\min}$ by computing ε and σ from the formulas

[36]In general, for solutions with r^{-n} asymptotic behavior at infinity, using an iterated boundary condition on the computational outer boundary gives greater accuracy than using a Dirichlet boundary condition (York and Piran, 1982). For linear problems, for example, Cantor (1983) has proved that a sequence of solutions with the iterated boundary condition and with increasing $(\rho_{\max},z_{\max})$ will converge to the true solution with $(\rho_{\max},z_{\max})=(\infty,\infty)$, and this sequence can even be used to study the asymptotic behavior of the true solution at $r=\infty$. Such strong statements cannot in general be made when a Dirichlet boundary condition is used on the outer boundary. In the Abelian Higgs model, where φ approaches its asymptotic value exponentially at infinity, the difference between the two types of boundary conditions is not expected to be as marked as in the case of power-law asymptotic behavior.

TABLE III. Node and half-node assignments and boundary conditions for the models of Sec. II.

Model	Defined on node lattice	Defined on half-node lattice	Boundary conditions (or remarks) $\rho = 0$	$z = 0$	Outer boundary $\rho = \rho_{\max}$ or $z = z_{\max}$
Abelian Higgs Sec. II.C	B^0 (cf. Appendix C; $A^0 = A_C^0 + B^0$)		(charge coordinate $z = a$ midway between nodes) B^0 iterated	$B^0 = 0$	B^0 iterated
	φ		φ iterated	φ iterated	φ iterated (could also use $\varphi = 1$)
Leading logarithm Sec. II.D	Φ		(charge coordinate $z = a$ on node) $\Phi = Q,\ z < a$; $\Phi = \frac{1}{2}Q,\ z = a$; $\Phi = 0,\ z > a$	Φ iterated	$\Phi = 0$
		σ	σ determined at all half-node points within computational mesh by updating from Φ		
Axially symmetric monopole Sec. II.E	h_1		h_1 iterated	$h_1 = 0$	$h_1 = (1 - n/r)\cos\vartheta$
	h_2		$h_2 = 0$	h_2 iterated	$h_2 = (1 - n/r)\sin\vartheta$
	f_1		$f_1 = n$	f_1 iterated	$f_1 = n\cos^2\vartheta$
	f_2		$f_2 = 0$	$f_2 = 0$	$f_2 = n\sin\vartheta\cos\vartheta$
	a_1		$a_1 = 0$	a_1 iterated	$a_1 = -(1/r)\sin\vartheta$
	a_2		a_2 iterated	$a_2 = 0$	$a_2 = (1/r)\cos\vartheta$

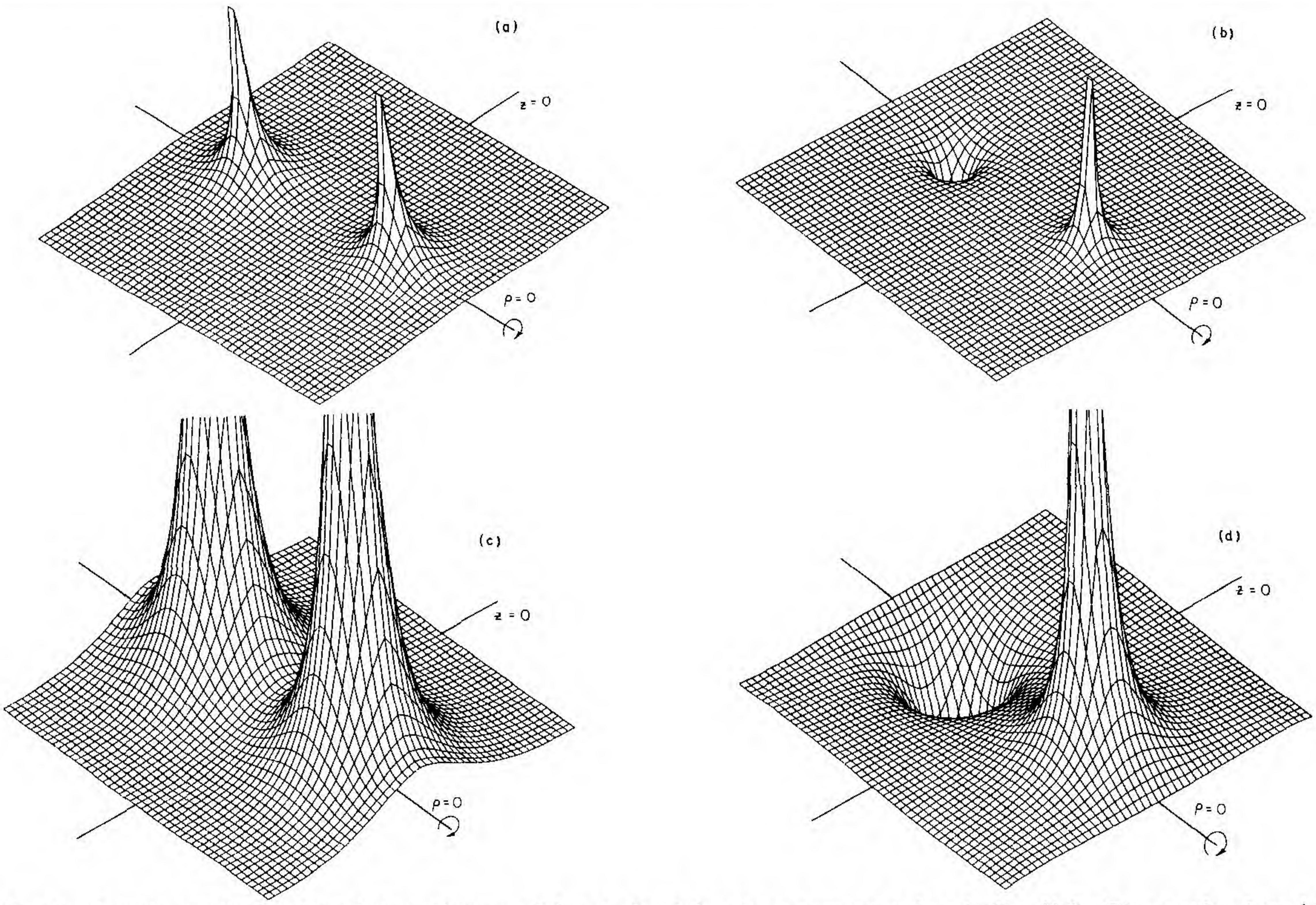


FIG. 12. Sample results for the Abelian Higgs model, calculated for the parameter values in Eq. (4.4). The graphs show (a) φ, (b) A^0, (c) φ with the vertical scale magnified by a factor of 10, and (d) A^0 with the same vertical scale magnification, all plotted vertically over a horizontal plane through the rotation axis. The values of φ and A^0 at the base of the figures are $\varphi = 1$ and $A^0 = 0$, respectively.

$$\varepsilon=\max[\tfrac{1}{4}b_0\log(E^2/\kappa^2),\varepsilon_{\min}]\ ,$$
$$\sigma=\min\left[\frac{2\pi\kappa}{\rho\,|\nabla\Phi|}f\left[\frac{|\nabla\Phi|}{\pi b_0\kappa\rho}\right],\frac{1}{\rho^2\varepsilon_{\min}}\right]. \tag{4.7}$$

This procedure avoids floating point underflows and overflows, and corresponds to keeping the differential equation for Φ just barely elliptic even outside the free boundary. Provided that $\varepsilon_{\min}$ is chosen to be very small (we have used values ranging from 10^{-15} to 10^{-35} with equally satisfactory results), the results for V_{static} are essentially independent of $\varepsilon_{\min}$. Although convergence of the iteration is improved by over-relaxation, we have found that the nonlinearity of the combined iteration of Eqs. (3.86a) and (3.86b) leads to instabilities in the free boundary shape, if one attempts to use ω values as large as the optimum ω appropriate to the linear subiteration of Eq. (3.86a). These instabilities are avoided by limiting ω to at most $\omega=1.7$ when iterating on meshes larger than 25×25. Full convergence within the free boundary requires about 1–2 min of CPU time on a VAX 11/780 computer for a 25×25 mesh, and around 1 h for a 100×100 mesh. Sample results on a 25×25 mesh [Adler and Piran (1982a)], computed for the parameter values

$$\kappa=1\ ,\quad Q=(\tfrac{4}{3})^{1/2},\quad b_0=9/(8\pi^2)\ ,$$
$$\rho_{\max}=z_{\max}=8,\quad \tfrac{1}{2}R=a=4\ , \tag{4.8}$$

are given in Figs. 13(a)–13(d). These figures show, respectively, the flux function Φ, the field energy density $\mathcal{H}$, $\mathcal{H}$ with a vertical scale magnification of 100, and the *logarithm* of the dielectric constant ε, all plotted vertically over a horizontal plane through the rotation axis. In the plot of Φ the $\rho=0$ boundary condition of Eq. (4.6) is clearly visible, and in both plots of $\mathcal{H}$ one can see the Coulomb energy peaks. The plot of Φ and the magnified plot of $\mathcal{H}$ show that the flux and energy are confined within an oval curve, approximating the continuum limit free boundary, which is also clearly visible in the contour plots of Φ and $\mathcal{H}$ shown in Figs. 14(a) and 14(b). Both because we have imposed a cutoff $\varepsilon>\varepsilon_{\min}=10^{-15}$, and because of finite mesh-spacing effects, the computational problem has low-level residual structure extending outside the continuum free boundary (but lying within a second, computational, free boundary), as can be seen in the plot of $\log\varepsilon$ in Fig. 13(d). This residual structure, together with the fact that the location of the free boundary is not stationary under small variations around the equilibrium

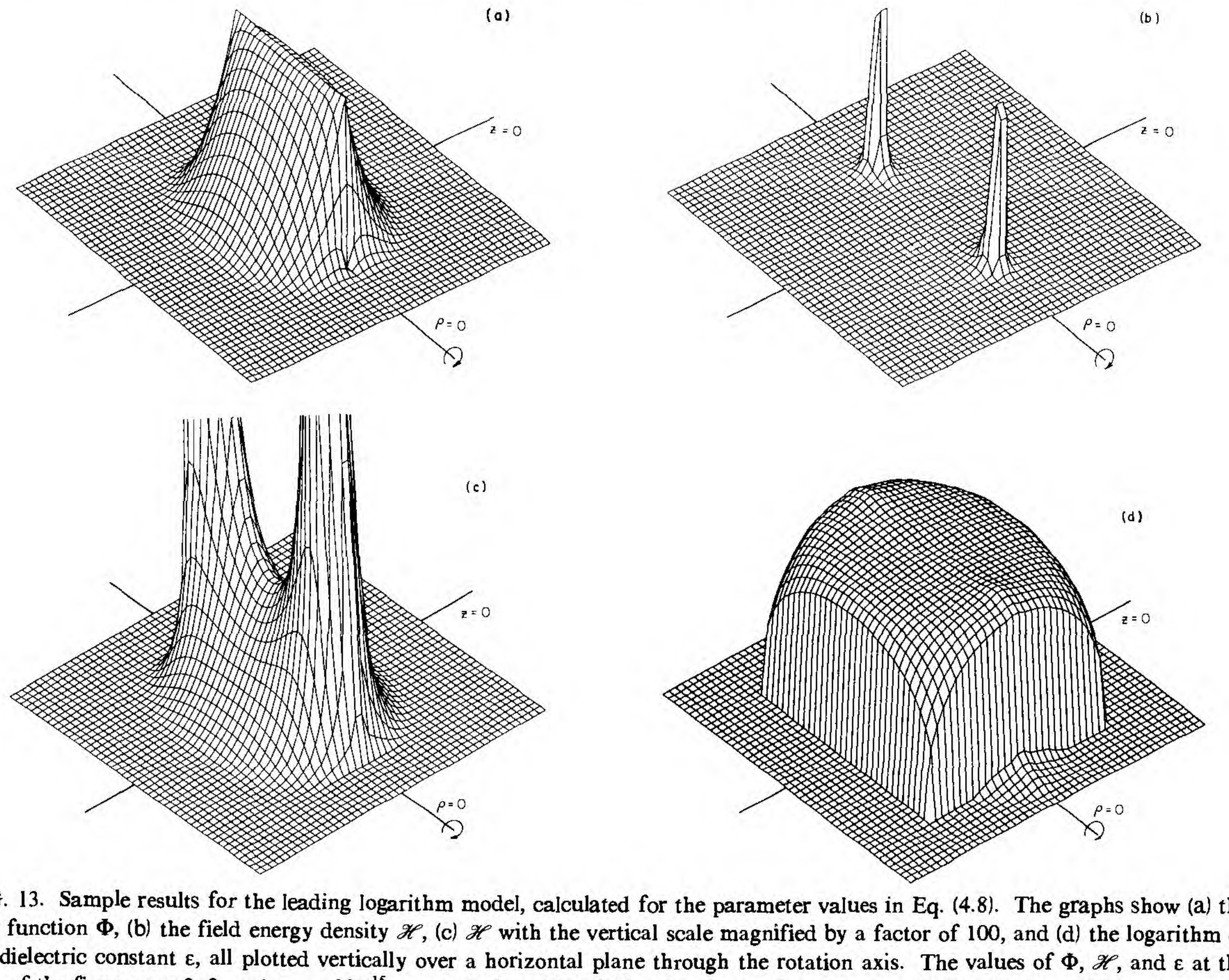


FIG. 13. Sample results for the leading logarithm model, calculated for the parameter values in Eq. (4.8). The graphs show (a) the flux function Φ, (b) the field energy density $\mathcal{H}$, (c) $\mathcal{H}$ with the vertical scale magnified by a factor of 100, and (d) the logarithm of the dielectric constant ε, all plotted vertically over a horizontal plane through the rotation axis. The values of Φ, $\mathcal{H}$, and ε at the base of the figures are 0, 0, and $\varepsilon_{\min}=10^{-15}$, respectively, with ε falling 14 decades from the top of (d) to the base. [When $\varepsilon_{\min}$ is reduced to 10^{-35}, the residual structure along the axis at the base of (d) is eliminated.]

solution Φ, makes an accurate determination of the free boundary location more difficult computationally than an accurate measurement of $V_{\rm static}$, which is stationary around equilibrium [cf. Eq. (2.42)]. An analytic investigation by Lehmann and Wu (1983), described briefly below, shows that in the limit as $R\to\infty$, the continuum free boundary approaches an ellipsoid of revolution. Rigorous proofs of the existence of a continuum free boundary have been given by Lieb (1983) and by Gidas and Caffarelli (1983).

As described above in Sec. III.E, to measure $V_{\rm static}(R)$ we make a sequence of measurements of $V_{\rm static}(R)-V_{\rm static}(R/2)$, with mesh geometries at the separations R, $R/2$ chosen so that the Coulomb self-energies cancel. For R values between 128 and 1, the calculation can be done in the original ρ,z coordinates, with uniform mesh spacings $\Delta\rho,\Delta z$ in the ρ and z directions. For R values smaller than 1, a Jacobian transformation as described in Sec. III.D is necessary. A simple "stretching"-type transformation which gives good results down to the smallest R values is given by

$$\begin{aligned}
&\rho=H(\rho',0.8a)\,,\\
&z=H(z',a)\,,\\
&H(z',a)=\begin{cases} z', & z'\le 3a\\ \dfrac{3a}{\left[4-\dfrac{z'}{a}\right]^{1/3}}, & 3a\le z'<4a\,,\end{cases}
\end{aligned}\tag{4.9}$$

to be used with $z'_{\rm max}=4a,\rho'_{\rm max}=3.2a$. This transformation has continuous first derivatives, leaves the mesh uniform near the source charges (so that Coulomb self-energies still cancel), and has an outer region in which mesh points are distributed so as to sample in a roughly uniform way the field energy of a dipole source, as indicated by the following estimate,

$$\begin{aligned}
\mathscr{H}_{\rm dipole}d^3x &\propto \frac{(1+3\cos^2\vartheta)}{r^6}r^2dr\\
&=\begin{cases} \dfrac{4}{3}\left|d\left[\dfrac{1}{z^3}\right]\right|=0.05\dfrac{dz'}{a^4}\,, & \vartheta=0\\ \dfrac{1}{3}\left|d\left[\dfrac{1}{\rho^3}\right]\right|=0.03\dfrac{d\rho'}{a^4}\,, & \vartheta=\pi/2\,.\end{cases}
\end{aligned}\tag{4.10}$$

In the calculations for R values much smaller than 1 the free boundary cannot be resolved even on a 100×100 mesh, but at small separations the infrared contributions to the static potential are no longer dominant, and so there is no difficulty in making an accurate determination of $V_{\rm static}$ by using the transformation of Eq. (4.9). An alternative method for doing the calculation at short distances is to use the transformation to bispherical coordinates given in Eq. (3.63). [This transformation is in fact useful at large distances as well; in Fig. 14(c) we show a contour plot of the flux function Φ in bispherical coordi-

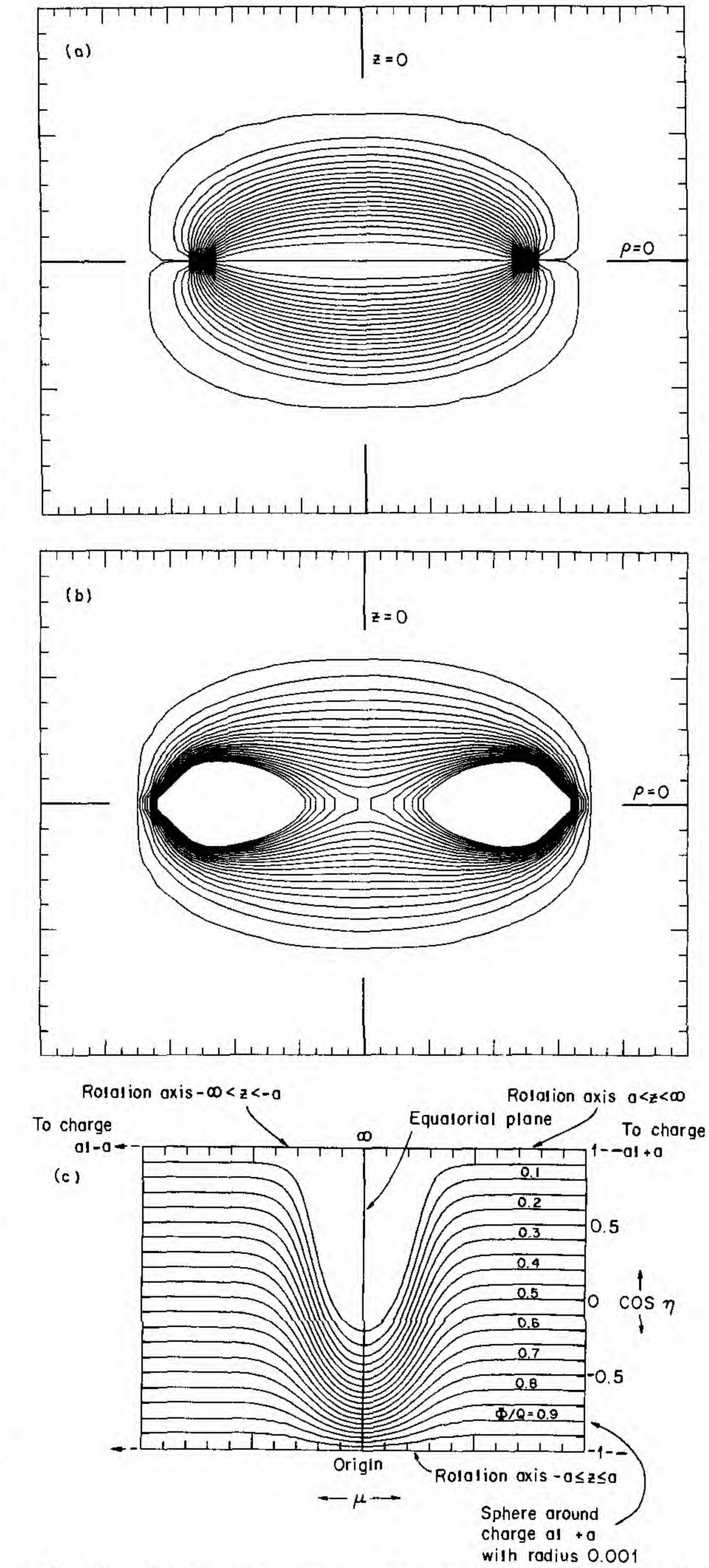


FIG. 14. Contour plots obtained from the numerical solution of the leading logarithm model, with the parameter values of Eq. (4.8). The plots (a) and (b) show contours of the flux function Φ and the energy density $\mathscr{H}$ in uniform Cartesian coordinates, corresponding to the elevation plots of Figs. 13(a) and 13(c), respectively. Specifically, (a) contains 21 equally spaced contours ranging from 0.001 to 1, while (b) contains 21 equally spaced contours ranging from 0.0001 to 0.01 (with the maximum of $\mathscr{H}$ scaled to 1). The plot (c) shows contours of the flux function Φ obtained on a bispherical grid. The coordinates are μ and $\cos\eta$, and the charges are located at $\mu=\pm\infty$. The free boundary is just inside the contour line $\Phi/Q=0.05$.

nates, calculated for the distance $a=4$ at which the free boundary is clearly resolved.] The results of the calculations using "stretched" cylindrical coordinates and using bispherical coordinates agree to better than 1% at all distances, and are presented in the form of a parametrized analytic fit to $V_{\rm static}$ in Adler and Piran (1982b).

In solving a complicated problem numerically, it is important to check the computer program, wherever possible, against analytic expressions which are available in limiting cases. In the case of the leading logarithm model, systematic analytic approximations can be developed in the small-R limit (Adler, 1983) and in the large-R limit (Lehmann and Wu, 1983). At small R a perturbation analysis in powers of an appropriately defined running coupling $\zeta(R)$ gives

$$V_{\rm static}(R) \underset{R\to 0}{=} \frac{-Q^2}{4\pi R\frac{1}{2}b_0}[\zeta(R)+O(\zeta^3)]\ ,$$

$$\zeta(R)=\frac{f(w_R)}{w_R}$$

$$=\frac{1}{\log w_R}+O\left[\frac{\log\log w_R}{(\log w_R)^2}\right]+O\left[\frac{1}{(\log w_R)^3}\right]\ , \tag{4.11}$$

$$w_R=\frac{1}{\Lambda_P^2R^2}\ ,\quad \Lambda_P=2.52\kappa^{1/2}\ .$$

When the numerical results for $V_{\rm static}$ in the range $\kappa^{1/2}R\sim 10^{-6}-10^{-8}$ are fit to the functional form of Eq. (4.11) with Λ_P adjustable, we find $\Lambda_P\approx 2.49$, in excellent agreement with the analytic result. At large R, a systematic expansion of the differential equation for Φ in powers of $1/R$ gives[37]

$$\Phi=\Phi^{(0)}(\rho/R^{1/2},z/R)+\frac{1}{R}\Phi^{(1)}(\rho/R^{1/2},z/R)+\cdots\ , \tag{4.12}$$

$$\Phi^{(0)}=Q\left[1-\tfrac{1}{2}\left[\frac{\pi b_0}{2Q}\right]^{1/2}\frac{a\rho^2\kappa^{1/2}}{a^2-z^2}\right]^2\ ,\quad a=\tfrac{1}{2}R\ ,$$

permitting the determination[37] of the leading two terms in the large-distance behavior of the static potential,

$$V_{\rm static}(R) \underset{R\to\infty}{=} \kappa QR+Q^{3/2}\frac{2}{3}\left[\frac{2}{\pi b_0}\right]^{1/2}\kappa^{1/2}\log(\kappa^{1/2}R)+\cdots\ . \tag{4.13}$$

This formula shows that the large R bound on the linear potential derived by Adler (1981a) is saturated. The numerical results for $V_{\rm static}$ in the range $R\sim 10-100$ yield coefficients of the R and $\log(\kappa^{1/2}R)$ terms which agree with the analytic results of Eq. (4.13) to better than 1%. According to Eq. (4.12), at large R the limiting behavior of the free boundary is an ellipsoid of revolution

$$1=\frac{z^2}{a^2}+\tfrac{1}{2}\left[\frac{\pi b_0}{2Q}\right]^{1/2}\left[\frac{\rho\kappa^{1/2}}{a^{1/2}}\right]^2\ , \tag{4.14}$$

with the major axis along the axis of rotation growing as R, and with the minor axis growing as $R^{1/2}$. A study of the structure of the free boundary using the numerical solutions for $R\sim 50-10^3$ shows that the outer contours of Φ have a shape agreeing well with this formula. Hence in both limiting cases in which analytic approximations are available, they are in excellent agreement with the results of the numerical solution for Φ.

D. Axially symmetric monopoles

In solving numerically for the axially symmetric monopoles, we use the leading terms of the asymptotic formulas of Eq. (2.99) as Dirichlet boundary conditions on the outer boundary, without including the $l=2$ or higher terms in the expansion. Consequently, the potentials obtained computationally will contain errors of order $1/\rho_{\max}^3, 1/z_{\max}^3$, which can be made small by choosing $\rho_{\max}$ and $z_{\max}$ large enough. An important check on convergence is to verify that the bound

$$H=4\pi\kappa\,|\,n\,| \tag{4.15}$$

is attained. Since the leading terms which are retained in Eq. (2.99) make a contribution to the energy density given by

$$\mathscr{H}_{\rm asymptotic}=\frac{n^2}{(\rho^2+z^2)^2}\ , \tag{4.16}$$

we must include an analytic correction for the energy lying outside the boundary of the computational mesh when we test Eq. (4.15). Following the notation of Eq. (3.70), we do this by writing

$$H=H_{\rm inside}+H_{\rm outside}\ , \tag{4.17a}$$

with

$$H_{\rm inside}=4\pi\int_0^{\rho_{\max}}\rho d\rho\int_0^{z_{\max}}dz\mathscr{H} \tag{4.17b}$$

determined computationally, and with a simple integration giving

[37]The solution $\Phi^{(0)}$ positions the charges on the free boundary, reflecting the fact that the distance between the charges and the free boundary (z_A-a in Fig. 6) vanishes relative to R as $R\to\infty$. The derivation of the logarithmic coefficient in Eq. (4.13) assumes the stronger statement that $(z_A-a)/R$ vanishes faster than $R^{-1}\log R$ as $R\to\infty$. The agreement of Eq. (4.13) with the numerical results gives *a posteriori* evidence for the validity of this assumption; for an analytic investigation of this issue see Lehmann and Wu (1983).

Note: In Eq. (4.14), $\kappa^{1/2}$ should read $\kappa^{1/4}$.

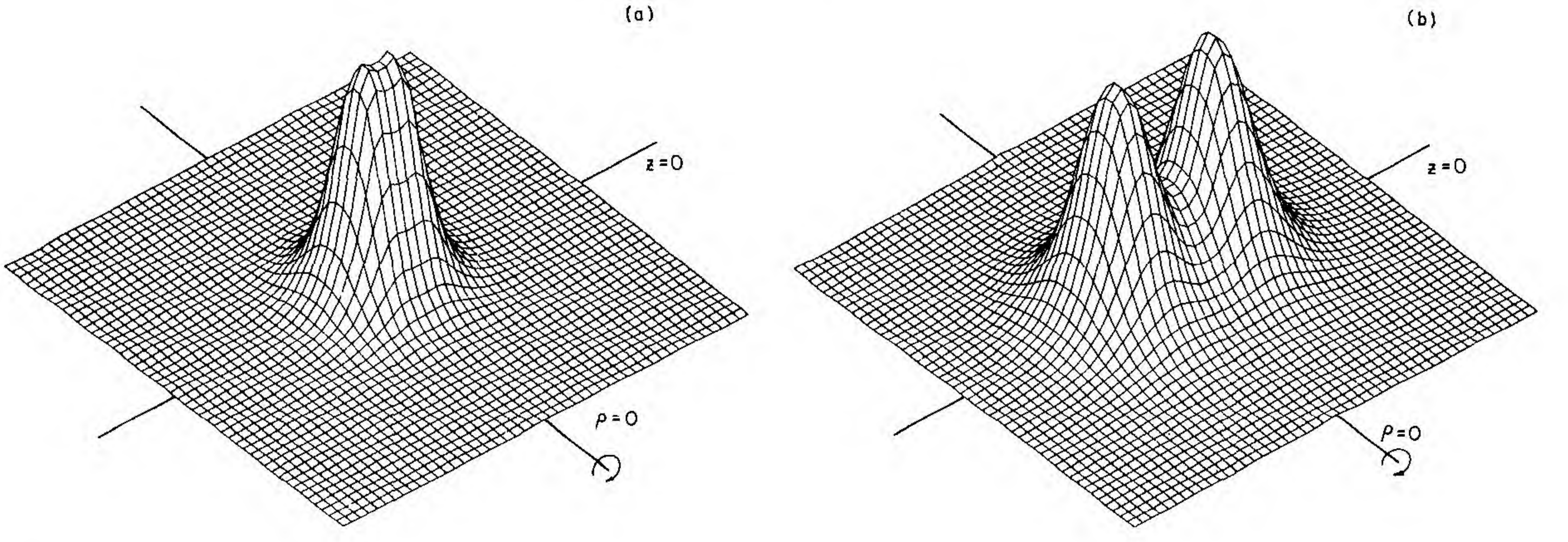


FIG. 15. Sample results for the axislly symmetric Bogomol'nyi-Prasad-Sommerfield monopole problem. The graphs show the field energy density $\mathscr{H}$ (a) for $n=1$ and (b) for $n=2$, plotted vertically over a horizontal plane through the rotation axis. The two cases can be characterized, respectively, as a fuzzy "ball" and a fuzzy "doughnut" of energy.

$$H_{\text{outside}}=\int_{\text{outside}}\mathscr{H}\approx\int_{\text{outside}}\mathscr{H}_{\text{asymptotic}}$$
$$=n^2\left[\frac{2\pi}{z_{\max}}+\frac{2\pi}{\rho_{\max}}\sin^{-1}\left[\frac{z_{\max}}{(\rho_{\max}^2+z_{\max}^2)^{1/2}}\right]\right]. \tag{4.17c}$$

As illustrations of the solution of the monopole equations, we have solved the $n=1,2$ cases on a mesh with $\rho_{\max}=z_{\max}=10$, in units in which $\kappa=1$ (and with the gauge function ψ of Appendix B taken as 0).[38] For the total energy H calculated from Eq. (4.17), we get (on the relatively coarse mesh plotted in Fig. 15)

$$\begin{aligned} &n=1,\quad H=12.88\,,\\ &n=2,\quad H=25.22\,, \end{aligned} \tag{4.18}$$

in good agreement with the values of 12.57, 25.13 expected from Eq. (4.15). In Fig. 15 we give plots of the energy density $\mathscr{H}$ for the solutions. The $n=1$ solution shows, in the large, the expected spherical symmetry, while the $n=2$ solution has the form of a fuzzy "doughnut," in agreement with a variational calculation of Rebbi and Rossi (1980) and with the recently found analytic solution for this case.[39] However, we also observe that the $n=1$ numerical solution has a dip in $\mathscr{H}$ along the rotation axis, whereas in the exact $n=1$ monopole solution given in Eq. (2.96) above, $\mathscr{H}$ decreases monotonically away from the axis. This dip is a remnant of the second-order errors in the discretization procedure which, in the present instance, weight too heavily the lattice cells along the axis, and thus lead to a reduction in the $\mathscr{H}$ value of the converged discrete solution at the axis. When the lattice spacing is made finer, the dip becomes smaller, but the correct monotinicity properties of $\mathscr{H}$ at $\rho=0$ appear only in the limit of zero mesh spacing.

ACKNOWLEDGMENTS

We wish to thank many colleagues in Princeton and elsewhere for their helpful questions and comments, and to thank Valerie Nowak for her beautiful work in computer composition of the manuscript. This work was supported by the U.S. Department of Energy under Grant No. DE-AC02-76ERO2220 and the National Science Foundation under Grant No. PHY-82-17352. Tsvi Piran acknowledges a grant from the Revson foundation.

APPENDIX A: MATHEMATICAL FEATURES OF THE LEADING LOGARITHM MODEL

We discuss in this appendix a number of mathematical features of the leading logarithm model. We begin by analyzing the structure of the real characteristic (the free boundary) of the differential equation

[38]As a check on the computer code, we have verified that the energy H of the numerical solution is unchanged when a choice $\psi\neq 0$ is used.

[39]For a discussion of the axially symmetric solutions realizable within the six-function ansatz, see, for example, Prasad and Rossi (1981).

$$\nabla\cdot[\sigma(\rho,|\nabla\Phi|)\nabla\Phi]=0\,, \tag{A1}$$

with σ defined in Eqs. (2.59) and (2.60) of the text. Using the chain rule and dividing by σ, we see that Eq. (A1) becomes

$$\nabla^2\Phi+\frac{\partial\log\sigma}{\partial\log|\nabla\Phi|}\nabla(\log|\nabla\Phi|)\cdot\nabla\Phi+\frac{\partial\log\sigma}{\partial\rho}\partial_\rho\Phi=0\,. \tag{A2}$$

This can be further rewritten by substituting

$$\begin{aligned}\nabla^2&=\partial_\rho^2+\partial_z^2+\rho^{-1}\partial_\rho\,,\\ \nabla(\log|\nabla\Phi|)\cdot\nabla\Phi&=\frac{\partial_i\Phi\partial_j\Phi}{|\nabla\Phi|^2}(\partial_i\partial_j\Phi)\\ &=\hat{n}_i\hat{n}_j\partial_i\partial_j\Phi\\ &=\hat{n}_i\partial_i\hat{n}_j\partial_j\Phi-|\nabla\Phi|\hat{n}_i\hat{n}_j\partial_i\hat{n}_j\\ &=\partial_n^2\Phi\,,\end{aligned} \tag{A3}$$

with $\hat{n}$ the unit normal defined by

$$\hat{n}=\frac{\nabla\Phi}{|\nabla\Phi|},\quad \hat{n}\cdot\hat{n}=1\,, \tag{A4}$$

thus giving

$$\left[(\partial_\rho^2+\partial_z^2-\partial_n^2)+\left[1+\frac{\partial\log\sigma}{\partial\log|\nabla\Phi|}\right]\partial_n^2\right]\Phi+\left[1+\frac{\partial\log\sigma}{\partial\log\rho}\right]\rho^{-1}\partial_\rho\Phi=0\,. \tag{A5}$$

Now from Eq. (2.59) we have

$$\begin{aligned}\log\sigma&=\log(2\pi\kappa)-\log\rho-\log|\nabla\Phi|+\log f(w)\,,\\ w&=\frac{|\nabla\Phi|}{\pi b_0\kappa\rho}=\frac{\partial_n\Phi}{\pi b_0\kappa\rho}\,,\end{aligned} \tag{A6}$$

and so the derivatives of σ appearing in Eq. (A5) can be expressed in terms of $f(w)$,

$$\begin{aligned}1+\frac{\partial\log\sigma}{\partial\log|\nabla\Phi|}&=\frac{\partial\log f}{\partial\log|\nabla\Phi|}=\frac{wf'(w)}{f(w)}\,,\\ 1+\frac{\partial\log\sigma}{\partial\log\rho}&=\frac{\partial\log f}{\partial\log\rho}=-\frac{wf'(w)}{f(w)}\,,\end{aligned} \tag{A7}$$

yielding Eqs. (2.72) and (2.73) of the text,

$$\begin{aligned}&[(\partial_\rho^2+\partial_z^2-\partial_n^2)+\alpha\partial_n^2]\Phi-\alpha\rho^{-1}\partial_\rho\Phi=0\,,\\ &\alpha=\frac{wf'(w)}{f(w)}=\frac{w}{w+f(w)}\,.\end{aligned} \tag{A8}$$

In the vicinity of a general, off-axis point B with coordinates x_B on the free boundary, we have

$$-\alpha\rho^{-1}\partial_\rho\Phi\approx-w\rho^{-1}\partial_\rho\Phi\sim|\nabla\Phi|^2\,, \tag{A9}$$

so to first order in $|\nabla\Phi|$, Eq. (A8) can be approximated by

$$\left[\partial_l^2+\frac{\partial_n\Phi}{\pi b_0\kappa\rho_B}\partial_n^2\right]\Phi=0\,, \tag{A10}$$

with l and n, respectively, tangential and normal Cartesian coordinates at the free boundary, as shown in Fig. 7. When the radius of curvature of the free boundary (and of the nearby level surfaces of Φ) is R_B, then to the needed accuracy the behavior of Φ near the free boundary has the form

$$\Phi=F\left[n-\tfrac{1}{2}\frac{l^2}{R_B}\right],\quad F(0)=0\,. \tag{A11}$$

Substituting Eq. (A11) into Eq. (A10) determines the function $F(z)$ to be

$$F(z)=\tfrac{1}{2}\frac{\pi b_0\kappa\rho_B}{R_B}z^2\,, \tag{A12}$$

giving Eq. (2.75) of the text.

Let us consider next the point $\rho=0,z=z_A$, where the free boundary intersects the axis of rotation. From Eq. (2.62) we know that $A^0=+\infty$ at the source charge Q, and A^0 is arbitrarily large within a sufficiently small neighborhood of the point $\rho=0,z=a$. On the other hand, since A^0 can be determined along the free boundary by integrating $\partial_l A^0=\kappa$ out from the plane $z=0$, where A^0 vanishes, we have

$$A^0(\rho=0,z=z_A)=\kappa L\,, \tag{A13}$$

with L the (finite) length of the segment of the free boundary lying within the quadrant drawn in Fig. 6. Hence A^0 is finite at $\rho=0,z=z_A$, and so we must have $z_A>a$, with the possibility $z_A=a$ excluded.

Let us suppose now that we have solved Eq. (2.56) or (A8), and hence know $\varepsilon=(\rho^2\sigma)^{-1}$ as a function of x. As mentioned in the text, one way (not the simplest way!) to determine A^0 is to solve the linear differential equation

$$\nabla\cdot(\varepsilon\nabla A^0)=-j^0 \tag{A14}$$

within the free boundary. Since Eq. (A14) is the Euler-Lagrange equation corresponding to minimization (for fixed $\varepsilon\geq 0$) of the functional

$$\int d^3x[\tfrac{1}{2}\varepsilon(\nabla A^0)^2-j^0A^0]\,, \tag{A15}$$

solutions will exist. To see that the solution is unique, even without the imposition of a boundary condition on the free boundary, let us suppose that Eq. (A14) has two C^1 solutions A_1^0 and A_2^0, so that $\delta A^0=A_1^0-A_2^0$ satisfies

$$\nabla\cdot(\varepsilon\nabla\delta A^0)=0\,. \tag{A16}$$

Multiplying by δA^0 and integrating over the interior of

the free boundary, we have

$$\begin{aligned}0&=\int d^3x\delta A^0\nabla\cdot(\varepsilon\nabla\delta A^0)\\&=\int d^3x\nabla\cdot(\delta A^0\varepsilon\nabla\delta A^0)-\int d^3x\varepsilon(\nabla\delta A^0)^2\\&=\int_{\text{free boundary}}d\mathbf{S}\cdot(\delta A^0\varepsilon\nabla\delta A^0)-\int d^3x\varepsilon(\nabla\delta A^0)^2\,.\end{aligned}\tag{A17}$$

The first term in the final line of Eq. (A17) vanishes, because ε vanishes on the free boundary, and so Eq. (A17) implies that $\nabla\delta A^0=0$ in the interior. This, together with the requirement that δA^0 be an odd function of z, implies the vanishing of δA^0 within the free boundary.

The seemingly paradoxical fact that Eq. (A5) requires a Dirichlet condition $\Phi=0$ on the free boundary, while Eq. (A14) requires no boundary condition for A^0 on the free boundary, has an interpretation in terms of the general boundary-value problem for second-order equations with non-negative characteristic form, given by Fichera (1956). The generalized Dirichlet problem takes the form[17]

$$\begin{aligned}&L(u)=a^{kj}(x)\partial_k\partial_j u+b^k(x)\partial_k u+c(x)u=f(x)\text{ in }\Omega\,,\\&u=g\text{ on }\Sigma_2\cup\Sigma_3\,,\end{aligned}\tag{A18}$$

with f and g functions defined on Ω and on $\Sigma_2\cup\Sigma_3$, respectively. The sets Σ_2 and Σ_3 are subsets of the boundary Σ of the domain Ω, specified as follows. Let n_k be the inward directed normal to the boundary. The set Σ_3 is defined to be the noncharacteristic part of the boundary, where $a^{kj}n_kn_j>0$. The characteristic part of the boundary, where $a^{kj}n_kn_j=0$, is divided into sets $\Sigma_0,\Sigma_1,\Sigma_2$, defined by

$$\begin{aligned}&b=0\text{ on }\Sigma_0\,,\\&b>0\text{ on }\Sigma_1\,,\\&b<0\text{ on }\Sigma_2\,,\\&b=n_k(b^k-\partial_j a^{kj})\,.\end{aligned}\tag{A19}$$

According to Eq. (A18), a Dirichlet boundary condition is required on Σ_2 and Σ_3, while no boundary condition is needed on the subsets Σ_0 and Σ_1 of the boundary.

Let us now analyze Eqs. (A8) and (A14) using this formalism. In discussing Eq. (A8) it suffices to use the approximate form of Eq. (A10), with l and n fixed Cartesian axes as in Fig. 7, giving

$$a^{ll}=1,\ a^{ln}=0,\ a^{nn}=\frac{\partial_n\Phi}{\pi b_0\kappa\rho_B},\ b^l=b^n=c=0\,,\tag{A20}$$

so that

$$b=-\frac{\partial}{\partial n}a^{nn}=-\frac{1}{\pi b_0\kappa\rho_B}\partial_n^2\Phi=-\frac{1}{R_B}<0\,.\tag{A21}$$

Thus for Eq. (A8) the free boundary is in Σ_2, and the imposition of a Dirichlet condition $\Phi=0$ on the free boundary is required. [Note that this condition, together with the discontinuity of Φ at the source charges given by Eq. (2.50), then implies that $\Phi=Q$ on the interior line segment $\rho=0$, $|z|<a$.] In Eq. (A14), we have

$$\begin{aligned}&a^{ij}=\varepsilon\delta^{ij},\ b^k=\partial_k\varepsilon,\ c=0\\&\Longrightarrow b^k-\partial_j a^{kj}=0\,,\end{aligned}\tag{A22}$$

so that the free boundary is in Σ_0. Hence no boundary condition for A^0 on the free boundary is needed when ε has been determined as a function of x by first solving the equation for Φ.

Suppose, on the other hand, that we attempt to solve the full nonlinear problem for A^0 given by Eqs. (2.41) and (2.46) directly, with ε not known *a priori*. These equations, when recast in the standard quasilinear form of Eq. (2.70), yield

$$\varepsilon_{ij}\partial_i\partial_j A^0=-j^0\,,\tag{A23a}$$

with ε_{ij} the field-strength dependent dielectric tensor

$$\varepsilon_{ij}=\delta_{ij}\tfrac14 b_0\log[(\nabla A^0)^2/\kappa^2]+\tfrac12 b_0\hat l_i\hat l_j\,.\tag{A23b}$$

The unit vector $\hat l_i$ is defined by

$$\hat l_i=\frac{\partial_i A^0}{|\nabla A^0|}\,,\tag{A24a}$$

and since

$$\hat l\cdot\nabla\Phi\propto\mathbf{D}\cdot\nabla\Phi\propto(\hat\phi\times\nabla\Phi)\cdot\nabla\Phi=0\,,\tag{A24b}$$

$\hat l$ is orthogonal to the unit vector $\hat n$ of Eq. (A4). Comparing Eq. (A23) with Eq. (A18), we see that

$$a^{ij}=\varepsilon_{ij},\ b^i=c=0\,,\tag{A25}$$

and so Eq. (A23) is elliptic in the interior region and degenerates on the characteristic, with

$$\begin{aligned}b&=-\partial_n[\tfrac14 b_0\log(E^2/\kappa^2)]-\tfrac12 b_0\hat n_i\partial_j(\hat l_i\hat l_j)\\&=-\tfrac12 b_0(E^{-1}\partial_n E+\hat n_i\partial_l\hat l_i)\,.\end{aligned}\tag{A26}$$

At a point B on the free boundary where the radius of curvature is R_B, we see from Fig. 7 that

$$\hat n_i\partial_l\hat l_i=\frac{1}{R_B}\,,\tag{A27a}$$

while from Eqs. (2.61), (2.75), and (A6) we get the leading behavior of E in the vicinity of the free boundary,

$$\begin{aligned}&E\approx\kappa(1+w)=\kappa\left[1+\frac{\partial_n\Phi}{\pi b_0\kappa\rho_B}\right]\approx\kappa\left[1+\frac{n}{R_B}\right]\\&\Longrightarrow E^{-1}\partial_n E=\frac{1}{R_B}\,,\end{aligned}\tag{A27b}$$

giving

$$b=-\frac{b_0}{R_B}<0\,.\tag{A28}$$

Hence according to the Fichera criterion of Eq. (A19), a Dirichlet boundary condition for A^0 is required at all points x_B on the free boundary. This boundary condition is implicitly available in the form

$$A^0(x_B)=\int_{z=0\text{ plane}}^{x_B} dl\,\partial_l A^0=\int_{z=0\text{ plane}}^{x_B} dl\kappa\,, \quad \text{(A29)}$$

with $\int dl$ a line integral along the characteristic, but since $A^0(x_B)$ thus becomes a function of the geometry of the free boundary (which is not known *a priori*), this condition is difficult to implement in a numerical calculation. An important advantage of the flux function reformulation is that it replaces Eq. (A29) by the explicit Dirichlet boundary condition $\Phi(x_B)=0$.

APPENDIX B: STRUCTURE AND PROPERTIES OF THE SIX-FUNCTION ANSATZ

Substituting the six-function ansatz of Eq. (2.97) into the field-potential relations of Eq. (2.80), we get the following expressions for the various field-strength components (with $\partial_z=\partial/\partial z, \partial_\rho=\partial/\partial\rho$):

$$\begin{aligned}
&(\mathscr{D}_j\varphi^a)\hat{z}^a\hat{z}^j=\partial_z h_1+a_1h_2,\quad B^{aj}\hat{z}^a\hat{z}^j=\rho^{-1}(\partial_\rho f_1+a_2f_2)\,,\\
&(\mathscr{D}_j\varphi^a)\hat{\rho}_n^a\hat{z}^j=\partial_z h_2-a_1h_1,\quad B^{aj}\hat{\rho}_n^a\hat{z}^j=\rho^{-1}(\partial_\rho f_2-a_2f_1)\,,\\
&(\mathscr{D}_j\varphi^a)\hat{z}^a\hat{\rho}^j=\partial_\rho h_1+a_2h_2,\quad B^{aj}\hat{z}^a\hat{\rho}^j=-\rho^{-1}(\partial_z f_1+a_1f_2)\,,\\
&(\mathscr{D}_j\varphi^a)\hat{\rho}_n^a\hat{\rho}^j=\partial_\rho h_2-a_2h_1,\quad B^{aj}\hat{\rho}_n^a\hat{\rho}^j=-\rho^{-1}(\partial_z f_2-a_1f_1)\,,\\
&(\mathscr{D}_j\varphi^a)\hat{\phi}_n^a\hat{\phi}^j=\rho^{-1}(h_2f_1-h_1f_2),\quad B^{aj}\hat{\phi}_n^a\hat{\phi}^j=\partial_\rho a_1-\partial_z a_2\,.
\end{aligned} \quad \text{(B1)}$$

Substituting these expressions into Eq. (2.86) gives a formula for the Hamiltonian H directly in terms of $h_{1,2},\ldots,$

$$\begin{aligned}
H&=4\pi\int_0^\infty \rho d\rho\int_0^\infty dz\mathscr{H}\,,\\
\mathscr{H}&=\tfrac{1}{2}[(\partial_z h_1+a_1h_2)^2+(\partial_z h_2-a_1h_1)^2+(\partial_\rho h_1+a_2h_2)^2+(\partial_\rho h_2-a_2h_1)^2]\\
&\quad+\frac{1}{2\rho^2}[(\partial_z f_1+a_1f_2)^2+(\partial_z f_2-a_1f_1)^2+(\partial_\rho f_1+a_2f_2)^2+(\partial_\rho f_2-a_2f_1)^2]\\
&\quad+\tfrac{1}{2}(\partial_z a_2-\partial_\rho a_1)^2+\frac{1}{2\rho^2}(h_1f_2-h_2f_1)^2\,.
\end{aligned} \quad \text{(B2)}$$

Equations (B1) and (B2) and the boundary conditions at $\rho=0$, $z=0$, and $r=\infty$ given in Eqs. (2.98) and (2.99) have a residual Abelian gauge invariance of the form

$$\begin{aligned}
&h_1\to h_1\cos\delta-h_2\sin\delta\,,\\
&h_2\to h_2\cos\delta+h_1\sin\delta\,,\\
&f_1\to f_1\cos\delta-f_2\sin\delta\,,\\
&f_2\to f_2\cos\delta+f_1\sin\delta\,,\\
&a_1\to a_1+\partial_z\delta\,,\\
&a_2\to a_2+\partial_\rho\delta\,,
\end{aligned} \quad \text{(B3)}$$

with δ a function of ρ,z satisfying

$$\delta=0 \text{ at } z=0,\quad \rho=0,\quad r\to\infty\,. \quad \text{(B4)}$$

Hence in order for H to have a unique minimum, it is necessary to add to it a gauge-fixing term

$$H_{\mathrm{gf}}=4\pi\int_0^\infty \rho d\rho\int_0^\infty dz\mathscr{H}_{\mathrm{gf}}\,. \quad \text{(B5)}$$

In the numerical work of Sec. IV.D, we shall use the following choice of gauge-fixing:

$$\mathscr{H}_{\mathrm{gf}}=\tfrac{1}{2}(\partial_z a_1+\partial_\rho a_2-\psi)^2\,, \quad \text{(B6)}$$

with ψ an arbitrary function which vanishes at the boundaries. Minimization of $H+H_{\mathrm{gf}}$ then picks out the member of the gauge-equivalence class of minima of H which satisfies

$$\partial_z a_1+\partial_\rho a_2-\psi=0\,. \quad \text{(B7)}$$

Since the differential equation for δ,

$$(\partial_z^2+\partial_\rho^2)\delta=\psi-(\partial_z a_1+\partial_\rho a_2),\quad 0<z<\infty,\quad 0<\rho<\infty\,, \quad \text{(B8)}$$

with the boundary conditions of Eq. (B4), gives a well-posed Dirichlet problem, the gauge condition of Eq. (B7) is always attainable and completely breaks the gauge degeneracy.

Adding Eq. (B6) to the kinetic term for $a_{1,2}$ in $\mathscr{H}$ gives

$$\begin{aligned}
&\tfrac{1}{2}(\partial_z a_2-\partial_\rho a_1)^2+\tfrac{1}{2}(\partial_z a_1+\partial_\rho a_2-\psi)^2\\
&\quad=\tfrac{1}{2}[(\partial_z a_1)^2+(\partial_\rho a_1)^2+(\partial_z a_2)^2+(\partial_\rho a_2)^2+\psi^2]\\
&\qquad-\psi(\partial_z a_1+\partial_\rho a_2)+\partial_z a_1\partial_\rho a_2-\partial_z a_2\partial_\rho a_1\,.
\end{aligned} \quad \text{(B9)}$$

Although the final term in Eq. (B9) is a total derivative, it does not vanish when integrated over a finite domain $0\le z\le z_{\max}, 0\le\rho\le\rho_{\max}$, and so should not be dropped in the numerical work.

As discussed in the text, the minima of H in the sector with topological quantum number n are self-dual or anti-self-dual gauge fields. In the self-dual case, where $-\mathscr{D}_j\varphi^a=B^{aj}$, the use of Eq. (B1) gives the differential equations

PHYSICAL REVIEW D VOLUME 23, NUMBER 12 15 JUNE 1981

Effective-action approach to mean-field non-Abelian statics, and a model for bag formation

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 10 November 1980; revised manuscript received 19 January 1981)

I propose a simple set of equations for mean-field non-Abelian statics with c-number sources, at general inverse temperature β, working from the Euclidean path-integral representation of the Hamiltonian partition function. The problem of finding the background-field configuration, and the mean-field potential, for point sources can be reduced to a classical differential equation problem involving a suitably defined thermal effective action functional. As an application I study the interaction of a pair of static classical sources coupled to a quantized SU(2) gauge field, using the simplified model defined by keeping only the leading-logarithm renormalization-group improvement to the local Euclidean action functional. I prove that the mean-field potential in this model grows at least linearly with the source separation, giving a simple model for bag formation. The use of these methods to construct a leading approximation to the $q\bar{q}$ binding problem in SU(3) quantum chromodynamics is discussed in two appendices. Appendix A describes the use of color-charge-algebra methods to generate an equivalent classical source problem, while Appendix B develops the properties of the transformation to a running coupling constant for which the one-loop renormalization group is exact. As a consistency check, in Appendix C I calculate the total mean-field ground-state energy, with source kinetic terms included, and show that it has the expected form.

I. EFFECTIVE-ACTION FORMALISM FOR NON-ABELIAN STATICS

I analyze in this paper the question of calculating the mean-field potential of classical point sources coupled to a quantized SU(2) gauge field, at zero and at finite temperature. This problem is of interest both in itself as a mathematical model, and because arguments based on the use of color-charge algebras suggest[1] that c-number source models should give a leading approximation to the problem of calculating the heavy quark-antiquark static potential in quantum chromodynamics.

My analysis proceeds from a field-theoretic generalization of the Euclidean (imaginary-time) version of Feynman's sum over histories. In potential scattering in one dimension, with Minkowski Lagrangian

$$L_M=\frac{1}{2}\left(\frac{dx}{dt}\right)^2-V(x)\,, \tag{1}$$

the Euclidean sum over histories reads

$$\langle x_f|e^{-\beta H}|x_i\rangle=N\int_{x_i}^{x_f}[dx]e^{-S}\,. \tag{2}$$

On the left-hand side of Eq. (2) $|x_i\rangle$ and $|x_f\rangle$ are position eigenstates and H is the Hamiltonian, while on the right-hand side N is a normalization constant and S is the Euclidean action

$$S=\int_0^\beta dt\left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2+V(x)\right], \tag{3}$$

and $\int[dx]$ denotes a functional integration over all paths $x(t)$ obeying the boundary conditions $x(0)=x_i$, $x(\beta)=x_f$. Setting $x_f=x_i$ and integrating over initial states gives a formula for the partition function,

$$\begin{aligned}\mathrm{Tr}(e^{-\beta H})&=\int dx_i\langle x_i|e^{-\beta H}|x_i\rangle\\&=N\int dx_i\int_{x_i}^{x_i}[dx]e^{-S}\,,\end{aligned} \tag{4}$$

where the paths in Eq. (4) now run from $x(0)=x_i$ back to $x(\beta)=x_i$. The generalization of Eq. (4) to a boson field theory containing spin-0 scalar fields and spin-1 gauge fields, denoted collectively by ϕ, can be written as

$$Z=\mathrm{Tr}(e^{-\beta H})=N\int d\phi_i\int_{\phi_i}^{\phi_i}[d\phi]e^{-S_E}\,. \tag{5}$$

On the left-hand side of Eq. (5), H is the Hamiltonian operator defined from the stress-energy tensor

$$H=\int d^3x\,T^{00}\,, \tag{6}$$

while on the right, S_E is the Euclidean action

$$S_E=\int_0^\beta dt\int d^3x\,\mathcal{L}_E\,, \tag{7}$$

obtained by continuing g_{00} from -1 to 1 in the generally covariant form of the Minkowski Lagrangian density[2]

$$\mathcal{L}_E=-\mathcal{L}_M^{\mathrm{gen\ cov}}|_{-1=g_{00}\to 1}\,. \tag{8}$$

The trace on the left is understood to be evaluated in any canonical gauge, where the Hilbert space contains only physical states, while the path integral on the right again extends over periodic paths, with $\phi(0)=\phi(\beta)=\phi_i$. The following observation makes the form of Eq. (5) intuitively plausible: For a field theory of scalars and spin-1 gauge

fields, the generally covariant Lagrangian density is linear in g_{00},

$$\mathcal{L}_M^{\text{gen cov}} = \mathcal{L}_{(0)} + \mathcal{L}_{(1)} g_{00}\,, \tag{9}$$

with $\mathcal{L}_{(0,1)}$ independent of g_{00}, whence from Eq. (8) we have

$$\begin{aligned}\mathcal{L}_M &= \mathcal{L}_{(0)} - \mathcal{L}_{(1)}\,,\\ \mathcal{L}_B &= -(\mathcal{L}_{(0)} + \mathcal{L}_{(1)})\,.\end{aligned} \tag{10}$$

But forming the Minkowski energy density T^{00},

$$\begin{aligned}T^{00} &= g^{00}\mathcal{L}_M - 2\frac{\delta \mathcal{L}_M}{\delta g_{00}}\\ &= (-1)(\mathcal{L}_{(0)} - \mathcal{L}_{(1)}) - 2\mathcal{L}_{(1)} = \mathcal{L}_B\,,\end{aligned} \tag{11}$$

we see that it is identical to the Euclidean Lagrangian density $\mathcal{L}_B$. Hence, the Euclidean action of Eq. (7) is a functional representation of the operator βH, just as in the potential theory case. A detailed justification of Eq. (5) can be obtained by a transformation from the conventional canonical formalism given by Bernard.[3]

I now apply Eq. (5) to an SU(2) gauge theory (with gauge potential $\vec{b}_\mu$ and electric and magnetic fields $\vec{E}^j$ and $\vec{B}^j$) coupled to a system of massive sources, and replace the source current density by its expectation, represented by a time-independent c-number external source $\vec{j}_\mu$. The equilibrium gauge field can be studied by keeping only the terms in H and S_B which explicitly depend on the gauge field variables,[4] while omitting the source dynamics (hence, H in the following formulas is a truncated Hamiltonian, and not the Hamiltonian for a closed system). With this simplification, we have

$$Z[\vec{j}_\mu] = \mathrm{Tr}(e^{-\beta H}) = N\int d\vec{b}_{\mu i}\int_{\vec{b}_{\mu i}}^{\vec{b}_{\mu i}} d[\vec{b}_\mu]\, e^{-S_B}\,, \tag{12}$$

where on the left

$$H = \int d^3x \left(\frac{1}{g^2}\frac{1}{2}(\vec{E}^j\cdot\vec{E}^j + \vec{B}^j\cdot\vec{B}^j) - \vec{b}_\mu\cdot\vec{j}_\mu\right) \tag{13}$$

is an operator, while on the right

$$S_B = \int_0^\beta dt \int d^3x \left(\frac{1}{g^2}\frac{1}{2}(\vec{E}^j\cdot\vec{E}^j + \vec{B}^j\cdot\vec{B}^j) - \vec{b}_\mu\cdot\vec{j}_\mu\right) \tag{14}$$

is a functional. The mean-field potential[5,6] associated with the static external source distribution $\vec{j}_\mu = (\vec{j}_0 \neq 0, \vec{j}_i = 0)$, including self-energies, is defined as

$$\begin{aligned}\delta V_{\text{mean field}} &\equiv \left\langle \int d^3x\, \vec{b}_0\cdot\delta\vec{j}_0\right\rangle\\ &= \mathrm{Tr}\left(e^{-\beta H}\int d^3x\,\vec{b}_0\cdot\delta\vec{j}_0\right)\Big/\mathrm{Tr}(e^{-\beta H})\,.\end{aligned} \tag{15}$$

Since

$$\int d^3x\,\vec{b}_0\cdot\delta\vec{j}_0 = -\int d^3x\,\delta\vec{j}_0\cdot\frac{\delta}{\delta\vec{j}_0}H\,, \tag{16}$$

we can reexpress the mean-field potential directly in terms of the partition function[6]

$$\begin{aligned}\delta V_{\text{mean field}} &= \frac{1}{\beta}\int d^3x\,\delta\vec{j}_0\cdot\frac{\delta}{\delta\vec{j}_0}\ln Z[\vec{j}_\mu]\\ &= \delta\frac{1}{\beta}\ln Z[\vec{j}_\mu]\,,\end{aligned} \tag{17a}$$

$$\Rightarrow V_{\text{mean field}} = \frac{1}{\beta}\{\ln Z[\vec{j}_\mu] - \ln Z[\vec{0}]\}\,, \tag{17b}$$

where I have fixed the constant of integration so that $V_{\text{mean field}}$ vanishes for vanishing source density. The problem of calculating $Z[\vec{j}_\mu]$ can be further reexpressed in terms of a classical differential equation problem involving a classical background field $\vec{c}_\mu$ and a vacuum effective action functional $\Gamma[\vec{c}_\mu]$. To do this, we write

$$Z[\vec{j}_\mu] = e^{-\beta W[\vec{j}_\mu]}\,, \tag{18}$$

and we introduce the time-independent classical background field $\vec{c}_\mu(x)$ induced by the time-independent external source distribution $\vec{j}_\mu(x)$,

$$\begin{aligned}\vec{c}_\mu(x) &\equiv -\frac{\delta W[\vec{j}_\mu]}{\delta\vec{j}_\mu(x)}\\ &= Z^{-1}N\int d\vec{b}_{\mu i}\int_{\vec{b}_{\mu i}}^{\vec{b}_{\mu i}} d[\vec{b}_\mu]\left(\frac{1}{\beta}\int_0^\beta dt\,\vec{b}_\mu(x)\right)e^{-S_B}\,.\end{aligned} \tag{19}$$

In this notation, the mean-field potential is given by

$$V_{\text{mean field}} = -W[\vec{j}_\mu] + W[\vec{0}]\,. \tag{20}$$

Defining the Legendre-transformed functional $\Gamma[\vec{c}_\mu]$ by

$$W[\vec{j}_\mu] = \Gamma[\vec{c}_\mu] - \int d^3x\,\vec{c}_\mu(x)\cdot\vec{j}_\mu(x)\,, \tag{21}$$

a standard calculation[7] shows that

$$\frac{\delta\Gamma[\vec{c}_\mu]}{\delta\vec{c}_\mu(x)} = \vec{j}_\mu(x)\,. \tag{22}$$

Equations (18)–(22) are the principal result of this section. They show that the mean-field potential, for any inverse temperature β, can be calculated by solving the classical differential equation problem of minimizing the functional $\Gamma - \int d^3x\,\vec{c}_\mu\cdot\vec{j}_\mu$, with Γ the thermal effective action functional.[8,9] In the limit $\beta\to\infty$, where Γ reduces to the Euclidean vacuum effective action functional, this minimum problem reproduces the variational principle of the "Euclidean statics" method which I have advocated elsewhere,[1] but with some signifi-

cant differences in physical interpretation.[10]

According to Eqs. (28)–(22), the problem of studying the mechanism for confinement in the model discussed here can be rephrased in terms of the following two related questions.

(1) Is there a physically reasonable class of vacuum action functionals for which Eqs. (18)–(22) give a confining potential for static point sources? This question is answered in the affirmative in the following section.

(2) Does the exact vacuum action functional calculated from the functional integral of Eq. (20) belong to the confining class?

The methods appropriate to studying these questions are quite different. For a given functional or class of functionals Γ, the first question is one of classical analytic or numerical methods for investigating partial differential equations. In the following section, Eq. (22) is investigated analytically for the leading-logarithm approximation to the renormalization-group improved local effective action functional, for which Γ takes the simple form $\Gamma[\vec{c}_\mu]=\Gamma(\vec{E}^j\cdot\vec{E}^j+\vec{B}^j\cdot\vec{B}^j)$; numerical methods of solution applicable to this class of functionals are currently being developed.[1,11] The second question is probably best studied by numerical Monte Carlo methods for doing the functional integral. Since confinement is an infrared effect, it should suffice to establish the properties of Γ for slowly varying source currents $\vec{j}_\mu$ and background fields $\vec{c}_\mu$. In this case, appropriate lattice transcriptions of the functional integral of Eq. (19) may give quantitatively accurate estimates of the behavior of the continuum effective action.

II. A SIMPLE MODEL FOR BAG FORMATION

As an illustration of the formalism developed in Sec. I, I analyze the following simple model, obtained by keeping only the leading-logarithm renormalization-group improvement[9] to the local Euclidean action functional

$$\Gamma[\vec{c}_\mu]=\int d^3x(\mathcal{L}_{\rm eff}-\mathcal{L}_{\rm eff}^{\rm min})\,, \tag{23}$$

with

$$\begin{aligned}
\mathcal{L}_{\rm eff}&=\mathcal{L}_{\rm eff}(F^2)\\
&=\frac{1}{2}\,\frac{(\vec{E}^j\cdot\vec{E}^j+\vec{B}^j\cdot\vec{B}^j)}{g^2}\left[1+\tfrac{1}{4}b_0g^2\ln\left(\frac{\vec{E}^j\cdot\vec{E}^j+\vec{B}^j\cdot\vec{B}^j}{\mu^4}\right)\right]\\
&=\tfrac{1}{8}b_0F^2\ln[F^2/(e\kappa^2)]\,,\\
\mathcal{L}_{\rm eff}^{\rm min}&=\mathcal{L}_{\rm eff}(\kappa^2)=-\tfrac{1}{8}b_0\kappa^2\,,\\
\kappa^2&=\frac{\mu^4}{e}e^{-4/(b_0g^2)}\,,\quad b_0=\frac{1}{8\pi^2}\,\frac{11}{3}\,C_2[\mathrm{SU(2)}]=\frac{1}{4\pi^2}\,\frac{11}{3}\,,\\
\vec{E}^j&=-\frac{\partial}{\partial x^j}\vec{c}^{\,0}-\vec{c}^{\,j}\times\vec{c}^{\,0}\equiv-\mathfrak{D}_j\vec{c}^{\,0}\,,\\
\vec{B}^j&=\epsilon^{jkl}\left(\frac{\partial}{\partial x^k}\vec{c}^{\,l}+\frac{1}{2}\vec{c}^{\,k}\times\vec{c}^{\,l}\right),\quad F^2\equiv\vec{E}^j\cdot\vec{E}^j+\vec{B}^j\cdot\vec{B}^j\,.
\end{aligned}\tag{24}$$

As has been extensively discussed in the literature,[12] the minimum of $\mathcal{L}_{\rm eff}$ occurs at the nonzero field strength $F=\kappa$. The source density $\vec{j}_0$ is taken to be a pair of classical sources of equal magnitude,

$$\begin{aligned}
&\vec{j}_0=\vec{Q}_1\delta^3(x-x_1)+\vec{Q}_2\delta^3(x-x_2)\,,\\
&|x_1-x_2|=R\,,\quad |\vec{Q}_1|=|\vec{Q}_2|=Q\,.
\end{aligned}\tag{25}$$

In analyzing the model defined by Eqs. (18)–(25), I make the physically plausible technical assumption that it suffices to minimize over potentials $\vec{c}_\mu$ for which $\vec{E}^j\cdot\vec{E}^j$ is axially symmetric around the line joining the sources.

The variational equations following from Eqs. (22)–(24) are

$$\mathfrak{D}_j(\epsilon\vec{E}^j)=\vec{j}_0\,, \tag{26a}$$

$$\epsilon^{kjm}\mathfrak{D}_j(\epsilon\vec{B}^m)=\vec{c}^{\,0}\times(\epsilon\vec{E}^k)\,, \tag{26b}$$

$$\epsilon=\epsilon(F^2)=\frac{\partial\mathcal{L}_{\rm eff}}{\partial(\frac{1}{2}F^2)}=\tfrac{1}{4}b_0\ln(F^2/\kappa^2)\,. \tag{26c}$$

Acting with $\mathfrak{D}_k$ on Eq. (26b) and using Eq. (26a) gives the constraint

$$0=\vec{c}^{\,0}\times\vec{j}_0\,; \tag{27}$$

hence, if we write

$$\vec{c}^{\,0}(x)=\hat{c}(x)c(x)\,,\quad \hat{c}\cdot\hat{c}=1\,, \tag{28}$$

we have

$$\hat{c}(x_1)\times\vec{Q}_1=\hat{c}(x_2)\times\vec{Q}_2=0\,, \tag{29}$$

which implies that

$$\vec{c}_0(x)\cdot\vec{j}_0(x)=c(x)j(x)\,, \tag{30}$$

with either

$$j(x)=\pm Q[\delta^3(x-x_1)-\delta^3(x-x_2)] \tag{31a}$$

or

$$j(x)=\pm Q[\delta^3(x-x_1)+\delta^3(x-x_2)]\,. \tag{31b}$$

From here on it will be convenient to work in the gauge with $\hat{c}(x)=\hat{z}$, in which we have

$$F^2=\left(\frac{\partial c}{\partial x^j}\right)^2+c^2(\vec{c}^{\,j}\times\hat{z})^2+\vec{B}^j\cdot\vec{B}^j\,. \tag{32}$$

I now will show that the minimization of W with respect to the vector potential $\vec{c}^{\,j}$ can be carried out explicitly, with the result

$$\min_{\vec{c}^{\,j}} W\equiv\bar{W}[c]=\int d^3x\left\{\bar{\mathcal{L}}_{\rm eff}\left[\left(\frac{\partial c}{\partial x^j}\right)^2\right]-c(x)j(x)\right\},$$

$$\bar{\mathcal{L}}_{\rm eff}\left[\left(\frac{\partial c}{\partial x^j}\right)^2\right]=0 \text{ for } \left(\frac{\partial c}{\partial x^j}\right)^2\le\kappa^2, \tag{33}$$

$$\bar{\mathcal{L}}_{\rm eff}\left[\left(\frac{\partial c}{\partial x^j}\right)^2\right]=\mathcal{L}_{\rm eff}\left[\left(\frac{\partial c}{\partial x^j}\right)^2\right]-\mathcal{L}_{\rm eff}^{\rm min} \text{ for } \left(\frac{\partial c}{\partial x^j}\right)^2\ge\kappa^2.$$

To prove Eq. (33), we note that $\mathcal{L}_{\rm eff}(F^2)-\mathcal{L}_{\rm eff}(\kappa^2)$ is a monotonic decreasing function of its argument for $0\le F^2\le\kappa^2$, and is a monotonic increasing function of its argument for $\kappa^2\le F^2$. Hence, W is minimized by the following choice of vector potential,

$$\begin{aligned}
&\vec{c}^{\,j}=\hat{z}\hat{z}^j a(\rho,z)\,,\quad \rho=(x^2+y^2)^{1/2}\,,\\
&a(\rho,z)=\int_0^\rho d\rho'\,A(\rho',z)\,,\\
&A(\rho',z)=0\,,\ \text{where } \left(\frac{\partial c}{\partial x^j}\right)^2\ge\kappa^2\,,\\
&A(\rho',z)=\left[\kappa^2-\left(\frac{\partial c}{\partial x^j}\right)^2\right]^{1/2},\ \text{where } \left(\frac{\partial c}{\partial x^j}\right)^2\le\kappa^2\,,
\end{aligned} \tag{34}$$

which gives (with $\hat{\phi}^j$ the azimuthal unit vector)

$$\begin{aligned}
&\vec{c}^{\,j}\times\hat{z}=0\,,\\
&\vec{B}^j=-\hat{z}\hat{\phi}^j A(\rho,z)\Rightarrow\\
&F^2=\max\left[\left(\frac{\partial c}{\partial x^j}\right)^2,\kappa^2\right],
\end{aligned} \tag{35}$$

and from which Eq. (33) immediately follows. (Note that it is at this point in the argument where the axial symmetry assumption has been used.) What is happening is that wherever the color-electric field is less than κ in magnitude, a color-magnetic field fills in to bring the total squared field strength up to the value κ^2 at which $\mathcal{L}_{\rm eff}$ is minimized.

We are now left with the purely Abelian problem of minimizing $\bar{W}[c]$, to which we apply simple flux conservation estimates introduced by 't Hooft.[13] Varying $\bar{W}$, we get the flux conservation equation

$$\begin{aligned}
&\frac{\partial}{\partial x^j}D^j=j(x)\,,\\
&D^j\equiv\bar{\epsilon}E^j\,,\quad E^j=-\frac{\partial}{\partial x^j}c\,,
\end{aligned} \tag{36}$$

with

$$\bar{\epsilon}=\begin{cases}\epsilon(E^jE^j) & \text{where } E^jE^j\ge\kappa^2\,,\\ 0 & \text{where } E^jE^j\le\kappa^2\,.\end{cases} \tag{37}$$

Evidently, wherever the E field strength is less than κ, the D field vanishes. This fact can be exploited to get a lower bound on $V_{\rm mean\,field}$ and an upper bound on $\bar{W}$, which by Eqs. (33), (36), and (37) can be rewritten as

$$\bar{W}_{\rm at\ equilibrium}=\int_{D>0}d^3x[\bar{\mathcal{L}}_{\rm eff}(E^jE^j)-E^jD^j]\,, \tag{38}$$

with the integral extending only over the region where D^j is nonvanishing. Dividing the integrand of Eq. (38) by $D=(D^jD^j)^{1/2}$, we get

$$D^{-1}[E^jD^j-\bar{\mathcal{L}}_{\rm eff}]=\kappa[\tfrac{1}{2}u+\tfrac{1}{2}f(u)] \tag{39}$$

with[14] (in the domain where $D>0$)

$$\begin{aligned}
&u=(E^jE^j)^{1/2}/\kappa\ge1\,,\\
&f(u)=\frac{u^2-1}{u\ln u^2}\ge1\,.
\end{aligned} \tag{40}$$

To turn Eq. (38) into a meaningful inequality, it is necessary to exclude the divergent self-energies of the charges by defining $\bar{W}^r$ to be the contribution to Eq. (38) coming from the exterior of small spheres of radius r centered on the charges. We then get

$$\begin{aligned}
&\bar{W}^r\le-\kappa I^r\,,\\
&V^r_{\rm mean\,field}\ge\kappa I^r\,,\\
&I^r=\int_\Lambda d^3x\,D\,,\quad \Lambda=\text{domain}\begin{cases}D>0\\ |x-x_1|\ge r\\ |x-x_2|\ge r\,.\end{cases}
\end{aligned} \tag{41}$$

The final step of the argument is to write $d^3x=dl\,dA$ with l the length along the flux lines of D^j and dA an element of area perpendicular to the flux lines. Denoting the flux by Φ, we have

$$\begin{aligned}
&dA\,D=d|\Phi|\ge d\Phi\,,\\
&\int_\Lambda d^3x\,D\ge\int_\Lambda d\Phi\,l(\Phi)\ge\Phi_{\rm tot}\,l_{\rm min}
\end{aligned} \tag{42}$$

with $l_{\rm min}$ the length of the shortest flux line. For the charge orientations of Eq. (31b), where the

flux lines terminate at infinity, we have

$$\Phi_{tot} = 2Q\,, \qquad l_{min} = \infty\,, \tag{43}$$

and I^r is infinite.[15] For the charge orientations of Eq. (31a), the flux lines run from the positive to the negative charge, giving

$$\begin{aligned} &\Phi_{tot} = Q\,, \\ &l_{min} = R - 2r\,, \\ &\tilde{W}^r \leq -\kappa Q(R-2r)\,, \\ &V^r_{\text{mean field}} \geq \kappa Q(R-2r)\,, \end{aligned} \tag{44}$$

which proves that the mean-field potential increases at least linearly for large R. In the limit of small R, a simple calculation shows that

$$V_{\text{mean field}} - \text{self-energies} \approx -\frac{Q^2}{4\pi R}\,\frac{g^2}{1+b_0 g^2 \ln(1/R\mu)}\,, \tag{45}$$

as expected from a leading-logarithm renormalization-group improved formalism. Hence, the simple model of Eqs. (18)–(25) interpolates smoothly between asymptotically free behavior at small source separations, and "baglike,"[16] confining behavior at large separations. The above analysis readily generalizes to the full renormalization-group improved local effective action functional,[9] provided that the effective action minimum remains at nonzero Euclidean field strength κ.[17] More generally, the results obtained above support the conjecture that a bag will form for large source separations, irrespective of the functional form of $\Gamma[\vec{c}_\mu]$, whenever the minimum of Γ occurs at potentials $\vec{c}_\mu$ with nonvanishing mean-square field strength.

ACKNOWLEDGMENTS

I wish to thank E. S. Fradkin, A. D. Linde, A. A. Migdal, A. B. Migdal, A. M. Polyakov, and A. E. Shabad for helpful comments and questions on the earlier version of these ideas contained in Ref. 1. I am greatly indebted to G. 't Hooft for a crucial conversation in which he explained the method of constructing flux conservation estimates[13] employed above in Eqs. (36)–(44). I also wish to thank M. S. Berger for helpful remarks about nonlinear systems, and R. F. Dashen and R. Jackiw for useful comments. This work was supported by the U. S. Department of Energy under Grant No. DE-AC02-76ER02220.

APPENDIX A: CONNECTION WITH SU(3) QUANTUM CHROMODYNAMICS

I briefly describe in this appendix how the methods of the text, together with the color-charge-algebra analysis of Ref. 1, can be applied to give a leading approximation to the $q\bar{q}$ binding problem in SU_3 quantum chromodynamics (QCD). In the static quark limit,[1] the gluon source current for the $q\bar{q}$ binding problem in QCD is

$$\begin{aligned} &j^{Aj} = 0\,, \\ &j^{A0} = Q_q^A \delta^3(x-x_1) + Q_{\bar{q}}^A \delta^3(x-x_2) \end{aligned} \tag{A1}$$

with Q_q^A and $Q_{\bar{q}}^A$ the quark and antiquark color-charge matrices. As discussed in Ref. 1, the gluon source current is now a 9×9 matrix operator acting on the nine-dimensional Hilbert space spanned by the $q, \bar{q}$ color states. The analysis of Sec. I can be extended to this case by including a factor $\frac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}}$ in all formulas and symmetrizing all inner products, so that Eq. (12) becomes

$$\begin{aligned} Z[j_\mu^A] &= \tfrac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}} \text{Tr}_{\text{gluon}}(e^{-\beta H}) \\ &= N \tfrac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}} \left(\int db^A_{\mu i} \int_{b^A_{\mu i}}^{b^A_{\mu f}} d[b^A_\mu] e^{-S_B} \right), \end{aligned} \tag{A2}$$

where on the left

$$\begin{aligned} H = \int d^3x \Big[&\frac{1}{g^2}\,\frac{1}{2}(E^{Aj}E^{Aj} + B^{Aj}B^{Aj}) \\ &- \frac{1}{2}(b^A_\mu j^A_\mu + j^A_\mu b^A_\mu) \Big] \end{aligned} \tag{A3}$$

is an operator in the product of the $q, \bar{q}$ and gluon Hilbert spaces, while on the right

$$\begin{aligned} S_B = \int_0^\beta dt \int d^3x \Big[&\frac{1}{g^2}\,\frac{1}{2}(E^{Aj}E^{Aj} + B^{Aj}B^{Aj}) \\ &- \frac{1}{2}(b^A_\mu j^A_\mu + j^A_\mu b^A_\mu) \Big] \end{aligned} \tag{A4}$$

is a functional in its dependence on the gluon variables, but is still a matrix operator in the finite dimensional $q, \bar{q}$ color Hilbert space.[18] Using cyclic invariance of the trace, the steps leading to Eqs. (18)–(22) go through just as before, giving

$$Z[j^A_\mu] = e^{-\beta W[j^A_\mu]}\,, \tag{A5}$$

$$V_{\text{mean field}} = -W[j^A_\mu] + W[0^A]\,, \tag{A6}$$

$$\delta W[j^A_\mu] \equiv -\tfrac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}} \left(\int d^3x\, c^A_\mu(x) \delta j^A_\mu(x) \right), \tag{A7}$$

$$W[j^A_\mu] = \Gamma[c^A_\mu] - \tfrac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}} \left(\int d^3x\, c^A_\mu(x) j^A_\mu(x) \right), \tag{A8}$$

$$\delta\Gamma[c^A_\mu] = \tfrac{1}{9}\,\text{tr}_q \text{tr}_{\bar{q}} \left(\int d^3x\, \delta c^A_\mu(x) j^A_\mu(x) \right). \tag{A9}$$

Note that Eq. (A7) defines c^A_μ to be a potential which, like the source current j^A_μ, is matrix valued in the nine-dimensional $q\bar{q}$ Hilbert space. To construct a QCD analog of the analysis of Sec. II, we must calculate a leading approximation to the effective action. In the classical limit, the effective action density is given by[19]

$$\mathcal{L}_{cl}=\frac{1}{2g^2}F^2,$$
$$F^2=\tfrac{1}{9}\,\mathrm{tr}_q\,\mathrm{tr}_{\bar q}(E^{Aj}E^{Aj}+B^{Aj}B^{Aj}),$$
$$E^{Aj}=-\frac{\partial}{\partial x^j}c^{A0}+iP_f^A(c^j,c^0),\tag{A10}$$
$$B^{Aj}=\epsilon^{jkl}\left(\frac{\partial}{\partial x^k}c^{Al}-\frac{i}{2}P_f^A(c^k,c^l)\right),$$
$$P_f^A(u,v)\equiv\frac{i}{2}f^{ABC}(u^Bv^C+v^Cu^B).$$

The renormalization-group improvement of Eq. (A10) is obtained by taking g^2 to be a running-coupling function of the argument $g^2\mathcal{L}_{cl}$ giving, in leading-logarithm approximation (cf. remarks in Ref. 28),

$$\Gamma[c_\mu^A]=\int d^3x[\mathcal{L}_{eff}(F^2)-\mathcal{L}_{eff}(\kappa^2)],\tag{A11}$$

$$\mathcal{L}_{eff}(F^2)=\tfrac{1}{8}b_0F^2\ln(F^2/e\kappa^2),\quad \kappa^2=\frac{\mu^4}{e}e^{-4/(b_0g^2)},$$
$$b_0=\frac{1}{8\pi^2}\frac{11}{3}C_2[SU(3)]=\frac{11}{8\pi^2}.\tag{A12}$$

To carry out the remainder of the analysis of Sec. II, we must reexpress Eqs. (A8)–(A12) in terms of number valued, as opposed to matrix valued, source density and gluon variables. To do this, let us recall[1] that the $q\bar q$ color-charge algebra is spanned by a basis $w_1^A,\ldots,w_4^A$, which satisfies the SU(2)×U(1) outer product algebra

$$P_f(w_r,w_s)=i\frac{1}{2}\epsilon_{rst}w_t,\quad r,s,t=1,2,3$$
$$P_f(w_r,w_4)=0,\tag{A13}$$

is orthonormal in the color-trace inner product,

$$\tfrac{1}{9}\,\mathrm{tr}_q\,\mathrm{tr}_{\bar q}(w_r^Aw_s^A)=\tfrac{8}{27}\delta_{rs},\tag{A14}$$

and over which the quark and antiquark color charges have the expansions

$$Q_q^A=\tfrac{3}{2}w_1^A+w_2^A+\frac{\sqrt5}{2}w_4^A,$$
$$Q_{\bar q}^A=\tfrac{3}{2}w_1^A-w_2^A-\frac{\sqrt5}{2}w_4^A.\tag{A15}$$

[As a check, we note that $\frac{1}{9}\mathrm{tr}_q\mathrm{tr}_{\bar q}(Q_q^AQ_q^A)=(8/27)\times(18/4)=4/3$.]

Expanding $Q_q^A, Q_{\bar q}^A, j_\mu^A, c_\mu^A, E^{Aj}$, and B^{Aj} over the basis w_r^A, with c-number coefficients, reduces the variational problem

$$\delta\Gamma[c_\mu^A]-\tfrac{1}{9}\,\mathrm{tr}_q\,\mathrm{tr}_{\bar q}\left(\int d^3x\,\delta c_\mu^A(x)j_\mu^A(x)\right)=0\tag{A16}$$

to a classical SU(2)×U(1) problem, analogous to that discussed in Sec. II. According to Eq. (A15), the U(1) effective q and $\bar q$ charges are opposite in sign, while the SU(2) effective charges have equal magnitudes. Hence, the quark and antiquark effective charges can be made antiparallel by an SU(2) gauge transformation,[20] leading to a solution with the same form as that obtained from Eq. (31a) in Sec. II, apart from the substitutions[21,22]

$$Q\to(\tfrac{4}{3})^{1/2},$$
$$b_0\to\frac{11}{8\pi^2}.\tag{A17}$$

APPENDIX B: TRANSFORMATION OF THE RUNNING COUPLING CONSTANT TO EXACT LEADING-LOGARITHM FORM

As already noted, the argument given in the text for a confining mean-field potential generalizes to the full renormalization-group improved local-effective-action functional, provided that the effective-action minimum remains at nonzero Euclidean field strength. When expressed in terms of the β function, this condition translates[9] into the requirement that the integral

$$\int^\infty\frac{dg'}{\beta(g')}\tag{B1}$$

should be convergent at its upper limit. Assuming convergence of the integral in Eq. (B1), I show in this appendix that one can make a nonanalytic transformation to a new running coupling constant g_R for which the one-loop renormalization-group structure is exact. The transformation is simply (with $\alpha_R=g_R^2$, $\alpha=g^2$)

$$\frac{1}{\alpha_R}=-\tfrac{1}{2}b_0\int_\alpha^\infty\frac{d\alpha'}{\bar\beta(\alpha')},\tag{B2}$$

where $\bar\beta=g\beta$ has the power-series expansion [with the coefficients given for SU(3) QCD with N_f light quark flavors[23]]

$$\bar\beta(\alpha)=-[\tfrac{1}{2}b_0\alpha^2+b_1\alpha^3+O(\alpha^4)],$$
$$b_0=\frac{1}{8\pi^2}(11-\tfrac{2}{3}N_f),\quad b_1=\frac{1}{2}\frac{1}{(8\pi^2)^2}(51-\tfrac{19}{3}N_f).\tag{B3}$$

Substituting the expansion of Eq. (B3) into Eq. (B2), we learn that for small running coupling constant,

$$\alpha_R=\alpha-a\alpha^2(\ln\alpha+\text{const})+\cdots,\quad a=\frac{2b_1}{b_0}\tag{B4}$$

and so $\alpha_R\to0$ when $\alpha\to0$. On the other hand, the convergence of Eq. (B1) implies that

$$\lim_{\alpha\to\infty}\int_\alpha^\infty\frac{d\alpha'}{\bar\beta(\alpha')}=0,\tag{B5}$$

and so $\alpha_R\to\infty$ as $\alpha\to\infty$. Hence the transformation of Eq. (B2) gives a nonsingular mapping from the half

(B2) gives a nonsingular mapping from the half line $0 \leq \alpha < \infty$ to the half line $0 \leq \alpha_R < \infty$. The renormalization-group structure in the new variable α_R is determined by $\bar{\beta}_R(\alpha_R)$, given by

$$\bar{\beta}_R(\alpha_R) = \bar{\beta}(\alpha)\frac{\partial \alpha_R}{\partial \alpha} = \bar{\beta}(\alpha)(-\alpha_R^2)\frac{\partial(\alpha_R^{-1})}{\partial \alpha}$$
$$= -\tfrac{1}{2} b_0 \alpha_R^2 , \tag{B6}$$

and so has exactly one-loop form.

A particularly interesting case of Eq. (B2) is obtained when $\bar{\beta}(\alpha)$ terminates at two-loop order,

$$\bar{\beta}(\alpha) = -(\tfrac{1}{2} b_0 \alpha^2 + b_1 \alpha^3), \tag{B7}$$

a situation which can always[24] be achieved [provided Eq. (B1) converges[25]] by an analytic transformation of the running coupling constant [i.e., by a rearrangement of the perturbation series which does not introduce coupling-constant logarithms]. In this case, Eq. (B2) can be explicitly integrated to give the transformation

$$\frac{1}{\alpha_R} = \frac{1}{\alpha} - a\left[\ln\left(\frac{1}{a\alpha}\right) + \ln(1 + a\alpha)\right], \tag{B8}$$

which for small $a\alpha$ can be developed into a series expansion

$$\frac{1}{\alpha_R} = \frac{1}{\alpha} - a\ln\left(\frac{1}{a\alpha}\right) + a\sum_{n=1}^{\infty}\frac{(-a\alpha)^n}{n} . \tag{B9}$$

The series of Eq. (B9) can be inverted by substituting

$$\alpha = \alpha_R(1 + \alpha_R f), \tag{B10}$$

which after some algebra gives

$$0 = \sum_{n=1}^{\infty}(-f)^n \alpha_R^{n-1} + a\ln(a\alpha_R) - a\sum_{n=1}^{\infty}\frac{(-\alpha_R f)^n}{n}$$
$$+ a\sum_{n=1}^{\infty}\frac{(-a\alpha_R)^n(1+\alpha_R f)^n}{n} . \tag{B11}$$

Substituting

$$f = \sum_{k=0}^{\infty}\alpha_R^k f_k \tag{B12}$$

into Eq. (B10), and equating the coefficients of like powers of α_R, gives explicit expressions for the coefficients f_k as polynomials in $\ln(a\alpha_R)$,

$$f_0 = a\ln(a\alpha_R),$$
$$f_1 = f_0^2 + af_0 - a^2, \ldots . \tag{B13}$$

Hence, starting from any convenient calculational scheme (for example, minimal subtraction in dimensional regularization), the QCD perturbation series can be reexpressed in terms of α_R by a two-step transformation: First, one transforms to a running coupling constant for which $\beta(\alpha)$ is given by Eq. (B7), and then one substitutes the inverse transformation given by Eqs. (B10)–(B13), yielding a series expansion in powers of α_R and $\ln(a\alpha_R)$.[26] In this series, the terms of order α_R^n contain only powers $0, 1, \ldots, n-1$ of $\ln(a\alpha_R)$.

An important property of the one-loop running coupling α_R is that it simultaneously maximizes the domains of analyticity of the renormalization-group improved local effective action $\mathcal{L}_{\rm eff}(F^2)$ and of the $\bar{\beta}$ function $\bar{\beta}(\alpha)$. For a general running coupling $\alpha(t)$, the renormalization-group improved local effective action density is given by[9]

$$\mathcal{L}_{\rm eff}(F^2) = \frac{1}{2}\frac{F^2}{\alpha(t)} , \tag{B14}$$
$$t = \tfrac{1}{4}\ln(F^2/e\kappa^2).$$

Substituting the one-loop running coupling α_R,

$$\frac{1}{\alpha_R(t)} = b_0 t \tag{B15}$$

gives

$$\mathcal{L}_{{\rm eff}\,R}(F^2) = \tfrac{1}{8} b_0 F^2 \ln(F^2/e\kappa^2), \tag{B16}$$

as used in Eq. (23) of the text. As a function of complex F^2, Eq. (B16) is analytic apart from a cut in the F^2 plane running along the negative real axis from $F^2 = 0$ to $F^2 = -\infty$. Such a timelike cut is expected from unitarity, and so $\mathcal{L}_{{\rm eff}\,R}$ has the maximum allowed analyticity domain in F^2. To study the analyticity properties of the general $\mathcal{L}_{\rm eff}(F^2)$, let us calculate the derivative

$$\frac{d}{d(F^2)}\mathcal{L}_{\rm eff}(F^2) = \frac{1}{2}\frac{1}{\alpha(t)} + \frac{1}{8}\frac{d}{dt}\left(\frac{1}{\alpha(t)}\right). \tag{B17}$$

From Eqs. (B2) and (B15), we get

$$\frac{d}{dt}\left(\frac{1}{\alpha(t)}\right) = \frac{\bar{\beta}(\alpha)}{-\frac{1}{2}\alpha^2} , \tag{B18}$$

and so

$$\frac{d}{d(F^2)}\mathcal{L}_{\rm eff}(F^2) = \frac{1}{2}\frac{1}{\alpha(t)} - \frac{1}{4}\frac{\bar{\beta}(\alpha(t))}{\alpha(t)^2} . \tag{B19}$$

We have seen above that at the spacelike F^2 where t vanishes, both α_R and α become infinite. Hence, $d\mathcal{L}_{\rm eff}(F^2)/d(F^2)$ is singular at spacelike F^2 unless $\bar{\beta}(\alpha)/\alpha^2$ is bounded as α becomes infinite. This is possible with $\bar{\beta}(\alpha)$ an entire function [which corresponds to the maximum allowed analyticity domain for the function $\bar{\beta}(\alpha)$] only if $\bar{\beta}(\alpha)/\alpha^2$ is a constant. Hence, the one-loop running coupling gives the maximal analytic extension of the renormalization-group substructure of QCD.[27] This result suggests that the one-loop model of Sec. II may give a universal, leading, semiclassical approximation to the confinement problem.[28]

APPENDIX C: TOTAL GROUND-STATE ENERGY

As a consistency check on the formalism of Sec. I, I show here that when source kinetic terms are included, the ground-state expectation of the *total* Hamiltonian for a system of two well-localized sources, in mean-field approximation, is

$$\langle 0|H_T|0\rangle = V_{\text{mean field}}(x_1, x_2) + \text{recoil terms} + \text{constant}. \tag{C1}$$

To most simply parallel the discussion of the text, I consider only the case of massive distinguishable fermion sources, with classical[29] SU_2 charges, for which H_T has the form

$$\begin{aligned} H_T &= \int d^3x\, T^{00} = \mathcal{H} + \mathcal{H}_{\text{kin}}, \\ \mathcal{H} &= \int d^3x \left(\frac{1}{g^2}\frac{1}{2}(\vec{E}^j\cdot\vec{E}^j + \vec{B}^j\cdot\vec{B}^j) - \vec{b}_0\cdot\vec{\mathcal{J}}_0\right), \\ \mathcal{J}_0 &= \psi_1^\dagger\vec{Q}_1\psi_1 + \psi_2^\dagger\vec{Q}_2\psi_2, \\ \mathcal{H}_{\text{kin}} &= \int d^3x(\psi_1^\dagger i\partial_0\psi_1 + \psi_2^\dagger i\partial_0\psi_2). \end{aligned} \tag{C2}$$

Taking the ground-state expectation of Eq. (C2), we have

$$\langle 0|H_T|0\rangle = \langle 0|\mathcal{H}|0\rangle + \langle 0|\mathcal{H}_{\text{kin}}|0\rangle. \tag{C3}$$

To apply mean-field theory, one assumes a Hartree factorization of the ground state (g = gluon, s = source)

$$|0\rangle = |0\rangle_g |0\rangle_s \tag{C4}$$

with

$$\begin{aligned} {}_g\langle 0|0\rangle_g &= {}_s\langle 0|0\rangle_s = 1, \\ {}_s\langle 0|\vec{\mathcal{J}}_0|0\rangle_s &= \vec{Q}_1\delta^3(x-x_1) + \vec{Q}_2\delta^3(x-x_2) = \vec{j}_0. \end{aligned} \tag{C5}$$

Hence, for $\langle 0|\mathcal{H}|0\rangle$ we get

$$\begin{aligned} \langle 0|\mathcal{H}|0\rangle &= {}_g\langle 0|H|0\rangle_g, \\ H &= \int d^3x\left(\frac{1}{g^2}\frac{1}{2}(\vec{E}^j\cdot\vec{E}^j + \vec{B}^j\cdot\vec{B}^j) - \vec{b}_0\cdot\vec{j}_0\right), \end{aligned} \tag{C6}$$

which involves the truncated Hamiltonian introduced in Sec. I. Since in the limit $\beta\to\infty$ only the ground state contributes to the partition function, from Eqs. (12), (18), and (20) of the text and Eq. (C6) we learn that

$$\langle 0|\mathcal{H}|0\rangle = W[\vec{j}_0] = -V_{\text{mean field}}(x_1, x_2) + \text{constant}. \tag{C7}$$

To evaluate the second term in Eq. (C3), we use the source field equations of motion

$$\begin{aligned} i\partial_0\psi_1 &= \vec{Q}_1\psi_1\cdot\vec{b}_0(x) + \text{recoil terms}, \\ i\partial_0\psi_2 &= \vec{Q}_2\psi_2\cdot\vec{b}_0(x) + \text{recoil terms}, \end{aligned} \tag{C8}$$

which, together with Eqs. (C4) and (C5), give

$$\langle 0|\mathcal{H}_{\text{kin}}|0\rangle = {}_g\langle 0|\int d^3x\,\vec{b}_0(x)\cdot\vec{j}_0(x)|0\rangle_g. \tag{C9}$$

The right-hand side of Eq. (C9) can be reexpressed in terms of the $\beta\to\infty$ limit of the partition function and then further rewritten using Eq. (19) of the text, giving

$${}_g\langle 0|\int d^3x\,\vec{b}_0\cdot\vec{j}_0(x)|0\rangle_g = \lim_{\beta\to\infty}\frac{1}{\beta}\left(\frac{\partial}{\partial\lambda}\ln Z[\lambda\vec{j}_0]\right)\bigg|_{\lambda=1} \tag{C10a}$$

$$= \int d^3x\,\vec{c}_0(x)\cdot\vec{j}_0(x). \tag{C10b}$$

Hence, we have

$$\langle 0|\mathcal{H}_{\text{kin}}|0\rangle = \vec{c}_0(x_1)\cdot\vec{Q}_1 + \vec{c}_0(x_2)\cdot\vec{Q}_2, \tag{C11}$$

and we can complete the proof of Eq. (C1) by showing that

$$\vec{c}_0(x_1)\cdot\vec{Q}_1 = V_{\text{mean field}}(x_1, x_2) + \text{constant}, \tag{C12a}$$

$$\vec{c}_0(x_2)\cdot\vec{Q}_2 = V_{\text{mean field}}(x_1, x_2) + \text{constant}. \tag{C12b}$$

To prove Eq. (C12a), let us write $\vec{c}_0(x)$ in the form

$$\vec{c}_0(x) = \vec{c}_0^{(A)}(x, x_1, x_2) + \vec{c}_0^{(B)}(x, x_1) \tag{C13}$$

with $\vec{c}_0^{(B)}$ chosen so that $\vec{c}_0^{(A)}$ is regular near $x = x_1$ and satisfies

$$[\delta_{x_1}\vec{c}_0^{(A)}(x, x_1, x_2)]|_{x=x_1} = 0. \tag{C14}$$

Using Eq. (15) of the text, in the $\beta\to\infty$ limit, we get

$$\begin{aligned} \delta_{x_1}V_{\text{mean field}}(x_1, x_2) &= \int d^3x\,\vec{c}_0(x)\cdot\vec{Q}_1\delta_{x_1}\delta^3(x-x_1) \\ &= \int d^3x\,\vec{c}_0^{(A)}(x, x_1, x_2)\cdot\vec{Q}_1\delta_{x_1}\delta^3(x-x_1) + \int d^3x\,\vec{c}_0^{(B)}(x, x_1)\cdot\vec{Q}_1\delta_{x_1}\delta^3(x-x_1). \end{aligned} \tag{C15}$$

The first term on the right of Eq. (C15) can be rewritten, by use of Eq. (C14), as

$$\delta_{x_1}\int d^3x\,\vec{c}_0^{(A)}(x, x_1, x_2)\cdot\vec{Q}_1\delta^3(x-x_1) - \int d^3x[\delta_{x_1}\vec{c}_0^{(A)}(x, x_1, x_2)]\cdot\vec{Q}_1\delta^3(x-x_1) = \delta_{x_1}[\vec{c}_0^{(A)}(x_1, x_1, x_2)\cdot\vec{Q}_1], \tag{C16}$$

while the second term on the right of Eq. (C15) is independent of x_2. Hence, Eq. (C15) implies

$$V_{\text{mean field}}(x_1, x_2) - \vec{Q}_1 \cdot \vec{c}_0(x_1) = V_1(x_1) + V_2(x_2), \quad \text{(C17)}$$

with V_1 independent of x_2 and V_2 independent of x_1. But since translational, rotational, and local SU_2 gauge invariance imply that both terms on the left-hand side of Eq. (C17) depend only on the relative distance $x_1 - x_2$, the terms V_1 and V_2 on the right must be constants, proving Eq. (C12a). A similar proof, using δ_{x_2}, gives Eq. (C12b).

According to Eqs. (C11) and (C12), a consistent mean-field approximation to the source wave equations is given by

$$\begin{aligned} i\partial_0\psi_1 &= \vec{c}_0(x_1)\cdot\vec{Q}_1\psi_1 + \text{recoil terms} \\ &= [V_{\text{mean field}}(x_1, x_2) + \text{constant}]\psi_1 + \text{recoil terms}, \\ i\partial_0\psi_2 &= \vec{c}_0(x_2)\cdot\vec{Q}_2\psi_2 + \text{recoil terms} \\ &= [V_{\text{mean field}}(x_1, x_2) + \text{constant}]\psi_2 + \text{recoil terms}. \end{aligned} \quad \text{(C18)}$$

These are just the usual one-body wave equations obtained from the potential theory of two sources, interacting through a potential $V_{\text{mean field}}(x_1, x_2)$, in the limit that the sources are well localized. Hence, the formalism of Sec. I reproduces all of the expected potential theory results.[30]

---

[1]R. Giles and L. McLerran, Phys. Lett. 79B, 447 (1978); S. L. Adler, Phys. Rev. D 17, 3212 (1978); S. L. Adler, *ibid.* 20, 3273 (1979). For a review, see S. L. Adler, in *The High Energy Limit*, proceedings of the of the 18th International School of Subnuclear Physics, "Ettore Majorana", edited by A. Zichichi (Plenum, New York, to be published). Further references are given here. [In QCD, the quantized gauge field is the underlying SU(3) gauge field, while the effective c-number sources lie in an unquantized, overlying SU(2) × U(1) gauge field. See Appendix A for a detailed discussion.]

[2]I define $\mathcal{L}_M^{\text{gen cov}}$ to be a scalar, so that

$$S_M = \int d^4x\,\sqrt{-g}\,\mathcal{L}_M^{\text{gen cov}}.$$

[3]C. W. Bernard, Phys. Rev. D 9, 3312 (1974). I thank L. Dolan for bringing this reference to my attention. For a discussion of the generalization of Eq. (5) to the case when fermions are present, see D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. 53, 43 (1981). Equations (12)–(14) remain valid for massive fermion sources at rest.

[4]The functional measure in Eq. (12) is understood to include the exponential of the gauge-fixing term, and the compensating Faddeev-Popov determinant (which can be represented as an additional functional integral over ghost fields). When the kinetic terms and functional integrals for the source fields producing $\vec{j}_\mu$ are included, S_g is properly gauge invariant, justifying use of the Faddeev-Popov functional measure. In the situation studied in this paper, where the only sources present are infinitely massive sources at rest, the source current can always be made time independent by an appropriate time-dependent gauge transformation. In such static-source gauges, the source functional integral can be omitted, leaving the expression for the partition function given in Eqs. (12)–(14), with $\vec{j}_\mu = (\vec{j}_0 \neq 0, \vec{j}_i = 0)$ and with $\vec{j}_0$ time independent. The static-source formalism is no longer invariant under all gauge transformations, but remains invariant under the subclass of time-independent gauge transformations. Since in a stationary state we have $d\langle\vec{b}_j\rangle/dt = d\vec{c}_j/dt = 0$, we are assured that the mean-field potential can be calculated from the expectation of the scalar potential $\langle\vec{b}_0\rangle = \vec{c}_0$.

[5]In a linear system the incremental potential $\delta V_{\text{mean field}}$ can be defined as either $\langle\int d^3x\,\vec{b}_0\cdot\delta\vec{j}_0\rangle$, which gives the mean energy change when an increment in source density $\delta\vec{j}_0$ is brought in from infinity, or as $\langle\delta\int d^3x\frac{1}{2}\vec{E}^j\cdot\vec{E}^j/g^2\rangle$, but in general these expressions are not equivalent: only the former can be used for nonlinear systems and is renormalization group invariant. When we study the nonrelativistic motion of the sources, the leading coupling of the sources to the gluon field involves only the values of $\vec{b}_0$ at the source positions. Hence, an average potential calculated from $\langle\int d^3x\,\vec{b}_0\cdot\delta\vec{j}_0\rangle$ is the correct starting point for a mean field, potential theory analysis of the source motion. See Appendix C for further details.

[6]From Eq. (17) we can see that the zero temperature ($\beta\to\infty$) mean-field potential is *not* the same as the static potential calculated from the Wilson loop, which in the notation used here is

$$V_{\text{static}} = \lim_{\beta\to\infty}\{W[i\vec{j}_\mu] - W[\vec{0}]\}.$$

The physical interpretation of V_{static} is that it is the ground-state eigenenergy of a static $q\bar{q}$ system. [See, for example, the derivation of the Wilson-loop formula given by L. S. Brown and W. I. Weisberger, Phys. Rev. D 20, 3239 (1979).] Since eigenenergies are defined only by continuation back to Minkowski space-time, it is not surprising that an imaginary source occurs when we formally represent V_{static} by a Euclidean path integral. The motivation for introducing $V_{\text{mean field}}$ is that it can be calculated strictly within the Euclidean formalism. In a perturbation expansion in the external source strength $\vec{j}_\mu$, the mean-field and Wilson-loop potentials agree in order $(\vec{j}_\mu)^2$, but differ beyond this order. In the Abelian case, there are no terms of higher order than $(\vec{j}_\mu)^2$, and so the two formalisms give the same static potential. In the non-Abelian case, the formalisms are inequivalent, and give different formulations of the confinement problem. It appears that the simple effective action approach to confinement developed in this paper can be obtained only by using a mean value formalism. I wish to thank R. F. Dashen for several discussions of these points.

(See also Appendix C and Ref. 30 below.)

[7]E. S. Abers and B. W. Lee, Phys. Rep. 9C, 1 (1973).

[8]Since Γ is not gauge invariant, the gauge-fixing conditions used in solving for $\bar{c}_\mu$ must be chosen to be compatible with the gauge noninvariance of Γ.

[9]The use of an effective action in this context was first suggested by H. Pagels and E. Tomboulis, Nucl. Phys. B143, 485 (1978).

[10]In particular, $\bar{c}_\mu$ is not to be used as Minkowski space Cauchy data and time evolved, as was implied in Ref. 1. In canonical gauges, the physical interpretation of $\bar{c}_\mu$ is that it is the expectation of $\bar{b}_\mu$, and is Minkowski time independent. Also, in Ref. 1 I used the incorrect, renormalization-group noninvariant formula for the potential (see Ref. 5 above).

[11]S. L. Adler and T. Piran, in *High Energy Physics—1980*, proceedings of the XXth International Conference, Madison, Wisconsin, edited by L. Durand and L. G. Pondrom (AIP, New York, 1981), p. 958.

[12]I. A. Batalin, S. G. Matinyan, and G. K. Savvidi, Yad. Fiz. 26, 407 (1977) [Sov. J. Nucl. Phys. 26, 214 (1977)]; G. K. Savvidy, Phys. Lett. 71B, 133 (1977); H. Pagels and E. Tomboulis, Ref. 9. See J. Ambjörn and P. Olesen, Nucl. Phys. B170, 60 (1980) for a discussion of corrections to the local effective action approximation, and extensive references. Many of these references consider only constant color fields, which has tended to obscure the fact that the gauge theory vacuum leading to the effective action of Eq. (23) is Lorentz invariant, with $\langle 0|\bar{b}^\mu|0\rangle = \langle 0|\vec{E}^j|0\rangle = \langle 0|\vec{B}^j|0\rangle = 0$ in the absence of sources. The vanishing of these expectations is reflected in the fact that the minimization of $\Gamma[\bar{c}_\mu]$ of Eq. (23) leads to a partially indeterminate variational problem, solved by any random color-electric and magnetic fields $\vec{E}^j$ and $\vec{B}^j$ satisfying $\vec{E}^j \cdot \vec{E}^j + \vec{B}^j \cdot \vec{B}^j = \kappa^2$. When sources are added, the variational problem of Eq. (22) is fully determinate only in the interior of the "bag", where $D > 0$, but remains partially indeterminate (in the sense described above) in the exterior region where $D = 0$. As a result, one cannot argue that there are nonvanishing gluon gauge potential or gauge field vacuum expectations by considering the limit of the exterior solution as a weak source is turned off; this line of reasoning applies only when the variational problem in the presence of sources is fully determinate in all of space.

[13]G. 't Hooft, in *Recent Progress in Lagrangian Field Theory and Applications*, proceedings of the Marseilles Colloquium, 1974, edited by C. P. Korthes-Altes *et al.* (Centre de Physique Theorique, Marseilles, 1975).

[14]To prove $f(u) \geq 1$, let $\psi = f - 1$, $\phi = 2/(u \ln u^2)$. A simple calculation shows that ψ satisfies the differential equation

$$\frac{d\psi}{du} + \phi\psi = \frac{(u-1)^2}{u^2 \ln u^2} > 0 \text{ for } u > 1.$$

Integrating up from $u = 1$ (where $\psi = 0$), this implies that ψ is positive for $u > 1$.

[15]In this case, one cannot neglect the surface term (as was done in the text) in the integration by parts leading from Eq. (33) to Eq. (38) but rather, one must work directly from Eq. (33). For the charge orientations of Eq. (31b), simple estimates (see H. Pagels and E. Tomboulis, Ref. 9) show that W has a positive infinite infrared divergence at equilibrium, corresponding to a vanishing partition function Z. Hence, the configuration with nonvanishing color flux at infinity is automatically excluded from the physical spectrum. Note that when the correspondence with QCD is made as in Appendix A, charge-conjugation symmetry or permutation symmetry will select either the effective charge orientations of Eq. (31a) or those of Eq. (31b), but not both. In the $q\bar{q}$ problem, the averaged potentials c_μ^A are charge-conjugation odd, selecting Eq. (31a). For the qq sector, the averaged potentials c_μ^A are symmetric under permutation of the sources, selecting Eq. (31b), and giving a vanishing partition function contribution. This is the expected result for a system which cannot be in a color singlet state.

[16]A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, Phys. Rev. D 9, 3471 (1974).

[17]If $\mathcal{L}_{\rm eff}$ attains its minimum at $F = 0$, and vanishes there as F^α, a simple estimate shows that the asymptotic behavior of the potential is $V_{\text{mean field}} \sim R^{(\alpha-3)/(\alpha-1)}$. This confines for $\alpha > 3$, but gives a linear potential only in the limit $\alpha \to \infty$. (See H. Pagels and E. Tomboulis, Ref. 9.)

[18]I am assuming a standard canonical quantization, in which only the constrained components of b_μ^A are matrix valued. Hence, the differentials $db_{\mu i}^A$ and $d[b_\mu^A]$ in Eq. (A2) are ordinary numbers. The assumption of canonical quantization is consistent with the conclusion reached at the end of the analysis, that the mean matrix-valued potential c_μ^A is Abelian apart from a time-independent gauge transformation. This means that c_0^A is nonzero, while c_j^A contains only a single spatial degree of freedom. The two spatial degrees of freedom in b_j^A which are orthogonal to c_j^A can then be canonically quantized by the standard Dirac bracket procedure. For example, taking the q and $\bar{q}$ to lie on the z axis, the gauge transformation rotating the $\bar{q}$ effective charge to be antiparallel to the q effective charge can be chosen to depend on z only, giving c_0^A, $c_z^A \neq 0$, but $c_{x,y}^A = 0$. This matrix-valued structure in the potentials is compatible with axial-gauge quantization.

[19]The classical limit of the effective action can be read off from Eq. (A2) by approximating e^{-S_B} by

$$e^{-S_B} \approx 1 - S_E,$$

and so is given by the field-strength terms in Eq. (A4), acted on by the quark color trace $\frac{1}{9}\,\mathrm{tr}_q\,\mathrm{tr}_{\bar{q}}$.

[20]Color-charge-algebra solutions of this form have been discussed by I. Bender, D. Gromes, and H. J. Rothe, Z. Phys. 5C, 151 (1980).

[21]In the formulation of Ref. 1, there arose the issue of how to fix the integration constants $K_{(t)}$ in the Lagrangian for the overlying algebra. The present analysis corresponds to taking the K_t's all equal, which differs from the rule which I had originally postulated.

[22]The effective Lagrangian analysis of the $q\bar{q}$ binding problem has the following Feynman diagram interpretation: Working in Coulomb gauge, the effective Lagrangian for the Coulomb gluons arises from Feynman diagrams which may be characterized as a central "blob," containing one or more closed gluon loops, from which $n \geq 2$ Coulomb gluons emerge. The effective Lagrangian contribution to $q\bar{q}$ binding is obtained

by stringing such "blobs" between q and $\bar{q}$ lines, attaching each emerging Coulomb gluon to either the q or the $\bar{q}$ line. This procedure yields the usual renormalization-group improved one-Coulomb-gluon exchange graph, and its nonlinear generalizations, which are responsible for the weak field-strength modification in the effective action which leads to confinement.

[23]Previously in this paper, I have taken N_f to be 0.

[24]G. 't Hooft, in *The Whys of Subnuclear Physics*, proceedings of the International School of Subnuclear Physics, Erice, 1977, edited by A. Zichichi (Plenum, New York, 1979), pp. 943–971. 't Hooft restricts his discussion to the case of analytic running coupling constant transformations. Nonanalytic transformations similar to those of Eq. (B8) have been recently investigated by Y. Frishman, R. Horsely, and U. Wolff, Phys. Lett. (to be published) and Weizmann Institute report (unpublished).

[25]N. N. Khuri and O. A. McBryan, Phys. Rev. D 20, 881 (1979).

[26]Similar coupling-constant logarithms have been found in three-dimensional QCD (which is related to the behavior of the four-dimensional theory at high-temperature phase transitions) by R. Jackiw and S. Templeton, Phys. Rev. D 23, 2291 (1981), and in chiral perturbation theory by H. Pagels, Phys. Rep. 16C, 219 (1975). In using the modified expansion to evaluate Euclidean Green's functions, it may be important to keep the $-i\epsilon$ in the Feynman denominators even after continuation to the Euclidean section. This gives a definite prescription for circling the spacelike pole in α_R and chooses a definite branch of the spacelike cut in $\ln\alpha_R$. The rearranged power series will in general contain imaginary contributions to the Euclidean Green's functions in each order, but (under the conventional assumption that the Euclidean Green's functions in QCD are real) these will cancel when the entire series is summed. Hopefully, the rearranged series will give real contributions to the Euclidean Green's functions which converge fast enough to give useful estimates (as, for example, is the case in the Wilson-Fisher expansion in critical phenomena when applied in 3 or 2 dimensions). Good convergence of the rearranged series would be an indication that the infrared behavior of QCD is effectively controlled by a weak coupling regime.

[27]There appears to be a close analogy between transformations of the radial coordinate in the theory of Schwarzschild black holes in general relativity, and transformations of the running coupling constant in QCD, with the concept of maximal analytic extension playing a key role in both cases. In both theories the natural coordinate (or coupling) in which one does calculations does not give the maximal analytic extension. Moreover, the transformations which yield the maximal extension have very similar functional form: Eq. (B8) relating α_R^{-1} to α^{-1} closely resembles the transformation $r^* = r + 2M\ln|r/2M - 1|$ which is used to remove the coordinate singularity at the horizon in black hole physics.

[28]A second interesting analogy is the fact that the leading logarithm effective action

$$\Gamma \propto \int d^4x F^2(x)\ln F^2(x)$$

has the same structure as the quantum-mechanical entropy

$$S = -k_B \operatorname{Tr}\rho\ln\rho,$$

which has many special and useful formal properties [see A. Wehrl, Rev. Mod. Phys. 50, 221 (1978)]. Perhaps this analogy can be exploited to understand the thermodynamic aspects of hadronic behavior. As a simple application of the entropy analogy, suppose that in the discussion of Appendix A we had applied the renormalization group improvement argument locally in the $q, \bar{q}$ color space, thus obtaining

$$l_{\text{eff}}(f^2) = \tfrac{1}{8}b_0 \operatorname{tr}_q \operatorname{tr}_{\bar{q}}[f^2\ln(f^2/e\kappa^2)],$$

$$f^2 = \tfrac{1}{3}(E^{Aj}E^{Aj} + B^{Aj}B^{Aj}),$$

instead of Eqs. (A10) and (A11), which in terms of f^2 read

$$\mathcal{L}_{\text{eff}}(F^2) = \tfrac{1}{8}b_0(\operatorname{tr}_q\operatorname{tr}_{\bar{q}} f^2)\ln(\operatorname{tr}_q\operatorname{tr}_{\bar{q}} f^2/e\kappa^2).$$

Since l_{eff} yields the same stress-energy tensor trace anomaly as does $\mathcal{L}_{\text{eff}}$, it is also an acceptable form for the effective action density. By some simple algebra, we find

$$l_{\text{eff}}(f^2) - \mathcal{L}_{\text{eff}}(F^2) = \tfrac{1}{8}b_0(\operatorname{tr}_q\operatorname{tr}_{\bar{q}} f^2)\operatorname{tr}_q\operatorname{tr}_{\bar{q}}(\rho\ln\rho),$$

$$\rho = f^2/(\operatorname{tr}_q\operatorname{tr}_{\bar{q}} f^2), \quad \operatorname{tr}_q\operatorname{tr}_{\bar{q}}\rho = 1.$$

Since ρ is a color density matrix, we can use the positivity of the entropy to conclude that

$$l_{\text{eff}}(f^2) \le \mathcal{L}_{\text{eff}}(F^2),$$

and so the use of l_{eff} would give at least as strong a linear potential as is obtained with $\mathcal{L}_{\text{eff}}$.

[29]The discussion of Appendix C is readily generalized to the QCD case by taking ${}_s\langle 0|\cdots|0\rangle_s$ to be an expectation with respect to the source spatial (but not color) wave functions, and including the source color wave functions in $|0\rangle_s$. Following Appendix A, the only changes are then the replacement of arrows by octet color indices, and the inclusion of a factor $\frac{1}{9}\operatorname{tr}_q\operatorname{tr}_{\bar{q}}$ in the inner products involving c_0 appearing in Eqs. (C10)–(C18).

[30]In contrast to the mean field approach, the Wilson-loop formula evaluates $\langle 0|H_T|0\rangle$ directly, without approximation, in terms of a Euclidean functional integral with imaginary sources. (See also the remarks in Ref. 6 above.)

Erratum: Effective-action approach to mean-field non-Abelian statics, and a model for bag formation [Phys. Rev. D 23, 2905 (1981)]

Stephen L. Adler

The discussion of Appendix C contains several errors. The conclusion that $\langle 0|\mathcal{H}_{\text{kin}}|0\rangle = 2V_{\text{mean field}}$ is correct, but in QCD this matrix element cannot be expressed in the form of Eq. (C11). Moreover, in general the decomposition of Eq. (C13) is not possible, and so the argument leading to Eq. (C12) is erroneous. A corrected version of Appendix C appears in S. Adler, in *Proceedings of the Fifth Johns Hopkins Workshop on Current Problems in Particle Theory*, edited by G. Domokos and S. K. Domokos (Johns Hopkins Univ., Baltimore, to be published).

In the sentence preceding the final paragraph of Sec. I, Eq. (20) should read Eq. (12).

Volume 113B, number 5 PHYSICS LETTERS 1 July 1982

FLUX CONFINEMENT IN THE LEADING LOGARITHM MODEL

Stephen L. ADLER and Tsvi PIRAN [1]
The Institute for Advanced Study, Princeton, NJ 08540, USA

Received 17 February 1982

We study the statics of quasi-abelian quark and antiquark source charges in the approximation in which leading logarithm radiative corrections are retained in the gauge gluon effective action functional. We show that the partial differential equation for the flux function is of degenerating elliptic type, leading to flux confinement within a free boundary which is a characteristic. The static potential increases linearly for large source separations, with a logarithmic subdominant term.

1. Introduction. As shown in a recent letter [1], by making a mean field approximation and a quark source charge approximation, the partition function for quantum chromodynamics (QCD) can be reduced to a relativistic model in which quarks couple to a pair of classical abelian background gauge fields obeying effective action dynamics. For the case of a single massive quark flavor, the functional integral formula for the S-matrix in this model is

$$S = \mathrm{ext}_{\hat{A}} \Big\{ \exp(\mathrm{i}\Gamma_{\mathrm{inv}}[\hat{A}])$$

$$\times \int \mathrm{d}[\bar{\psi}]\, \mathrm{d}[\psi]\, \exp\Big(\mathrm{i} \int \mathrm{d}^4 x\, \bar{\psi}(\mathrm{i}\hat{\not{D}} - m)\psi\Big)\Big\} ,$$

$$\hat{\not{D}} = \gamma^\mu \hat{D}_\mu , \quad \hat{D}_\mu = \partial_\mu - \mathrm{i}g\hat{A}^a_\mu \tfrac{1}{2}\hat{\lambda}^a ,$$

$$\hat{A}^{1,2,4,5,6,7}_\mu = 0, \quad \hat{A}^{3,8}_\mu \neq 0,$$

$$\hat{\lambda}^{1,2,4,5,6,7} = 0, \quad \hat{\lambda}^{3,8} = 2\lambda^{3,8} . \tag{1}$$

In eq. (1), the matrices $\hat{\lambda}^{3,8}$ are quasi-abelian quark effective charges, the functional $\Gamma_{\mathrm{inv}}[\hat{A}]$ is the gauge-invariant gluon [‡1] effective action, and the notation $\mathrm{ext}_{\hat{A}}\{\ \}$ indicates the extremum of the curly bracket over all values of the quasi-abelian gauge potentials $\hat{A}^{3,8}_\mu$. Making the standard rescaling by the coupling constant $\hat{A} \to \hat{A}/g$, introducing an effective action density by writing

$$\Gamma_{\mathrm{inv}}[\hat{A}] = \int \mathrm{d}^4 x\, \mathcal{L}_{\mathrm{eff}}[\hat{A}] , \tag{2}$$

and specializing to the case of static, infinitely massive sources, eq. (1) gives a formula for the static potential

$$S = \lim_{T\to\infty} \exp(-\mathrm{i}\, V_{\mathrm{static}} T)$$

$$= \lim_{T\to\infty} \mathrm{ext}_{\hat{A}} \Big\{ \exp\Big(\mathrm{i}T \int \mathrm{d}^3 x\, (\mathcal{L}_{\mathrm{eff}}[\hat{A}/g] - \hat{A}^a_0 \hat{J}^a_0$$

$$+ \text{mass terms})\Big)\Big\} ,$$

$$\hat{J}^a_0 = -\psi^\dagger \tfrac{1}{2}\hat{\lambda}^a \psi , \tag{3}$$

which can be rewritten (dropping the mass terms) as

$$V_{\mathrm{static}} = -\mathrm{ext}_{\hat{A}} \Big\{ \int \mathrm{d}^3 x\, (\mathcal{L}_{\mathrm{eff}}[\hat{A}/g] - \hat{A}^a_0 \hat{J}^a_0)\Big\} . \tag{4}$$

When solved using the classical approximation to the effective action,

$$\mathcal{L}_{\mathrm{eff}} \approx \tfrac{1}{2} g^{-2} \mathcal{F} , \quad \mathcal{F} \equiv -\tfrac{1}{2}(\partial_\mu \hat{A}^a_\nu - \partial_\nu \hat{A}^a_\mu)^2 , \tag{5}$$

eq. (4) gives classical abelian electrostatics. The *leading-logarithm model* is defined by including radiative cor-

[1] Also at Racah Institute for Physics, Hebrew University, Jerusalem, Israel.

[‡1] More generally, when there are light quarks which are observed only through their effect on the massive quark system, $\Gamma_{\mathrm{inv}}[\hat{A}]$ is the gauge-invariant effective action of the gluon plus light quark subsystem.

Volume 113B, number 5 PHYSICS LETTERS 1 July 1982

rections to $\mathcal{L}_{\mathrm{eff}}$ to leading logarithm order, giving

$$\mathcal{L}_{\mathrm{eff}} \approx \tfrac{1}{2} g^{-2} \mathcal{F}[1 + \tfrac{1}{4} b_0 g^2 \log(\mathcal{F}/\mu^4)], \tag{6}$$

where $b_0(>0)$ is the one-loop β-function confinement and μ is the subtraction point, $g^2 = g^2(\mu^2)$.

The model of eqs. (4) and (6) was studied for isolated quark sources by Pagels and Tomboulis [2], who showed that the infrared energy is linearly divergent. The leading logarithm model with a pair of oppositely charged point sources (as is appropriate to a q$\bar{\mathrm{q}}$ system),

$$\hat{j}_0^a = Q\hat{q}^a[\delta^3(x - x_1) - \delta^3(x - x_2)], \tag{7}$$

where $\hat{q}^a$ is the unit internal vector, was studied by Adler [3,4], who gave a variational argument [‡2] showing that for large source separations, V_{static} is bounded from below by a linear potential. In this letter we give the results of a detailed analytical and numerical study of the leading logarithm model, with particular emphasis on the structure of the domain within which the flux is confined.

2. Analytic formulation. Introducing scalar and vector potentials ϕ and A by writing

$$\hat{A}_0^a = \hat{q}^a\phi, \quad \hat{A}_j^a = \hat{q}^a A_j, \tag{8}$$

the variational problem of eqs. (4)–(8) can be rewritten as

$$V_{\mathrm{static}} = -\mathrm{ext}_{\phi,A}\left\{\int \mathrm{d}^3x\, [\mathcal{L}_{\mathrm{eff}}(\mathcal{F}) - \phi j_0]\right\},$$

$$\mathcal{F} = \boldsymbol{E}^2 - \boldsymbol{B}^2, \quad \boldsymbol{E} = -\nabla\phi, \quad \boldsymbol{B} = \nabla\times\boldsymbol{A},$$

$$j_0 = Q[\delta^3(x - x_1) - \delta^3(x - x_2)],$$

with $\mathcal{L}_{\mathrm{eff}}$ in the leading logarithm model given by

$$\mathcal{L}_{\mathrm{eff}}(\mathcal{F}) = \tfrac{1}{8} b_0 \mathcal{F} \log(\mathcal{F}/e\kappa^2),$$

$$\kappa^2 = (\mu^4/e)\exp[-4/(b_0 g^2)]. \tag{10}$$

The Euler–Lagrange and constraint equations implied by eq. (9) are

$$\nabla\cdot\boldsymbol{D} = j_0, \quad \nabla\times\boldsymbol{E} = 0,$$

$$\nabla\times\boldsymbol{H} = 0, \quad \nabla\cdot\boldsymbol{B} = 0,$$

$$\boldsymbol{D} = \epsilon\boldsymbol{E}, \quad \boldsymbol{H} = \epsilon\boldsymbol{B}, \quad \epsilon = \partial\mathcal{L}_{\mathrm{eff}}/\partial(\tfrac{1}{2}\mathcal{F}), \tag{11}$$

with the field-strength dependent dielectric constant ϵ given in the leading logarithm model by

$$\epsilon = \tfrac{1}{4} b_0 \log(\mathcal{F}/\kappa^2). \tag{12}$$

The source-free Euler–Lagrange equation for $\boldsymbol{H}$ can be satisfied by taking

$$\epsilon\boldsymbol{B} = 0, \tag{13}$$

giving two branches

$$\text{(I)} \quad \boldsymbol{B} = 0,$$

$$\text{(II)} \quad \epsilon = 0 \to \boldsymbol{B}^2 = \boldsymbol{E}^2 - \kappa^2. \tag{14}$$

Near the source charges, asymptotic freedom requires that the solution approach a Coulomb-like solution with $\boldsymbol{E}$ large and $\boldsymbol{B}$ vanishing; together with continuity, this implies that a finite domain containing the source charges lies on branch (I).

Specializing the analysis to branch (I), we are then left with a problem in nonlinear electrostatics,

$$\nabla\cdot\boldsymbol{D} = j_0, \quad \nabla\times\boldsymbol{E} = 0,$$

$$\boldsymbol{D} = \epsilon(E)\boldsymbol{E}, \quad \epsilon(E) = \partial\mathcal{L}_{\mathrm{eff}}(E^2)/\partial(\tfrac{1}{2}E^2). \tag{15}$$

Let us work henceforth in cylindrical coordinates $\rho = (x^2 + y^2)^{1/2}, z, \theta$, with the source charges located symmetrically on the z-axis at $z = \pm a$, and let $\hat{\boldsymbol{\theta}}$ be the azimuthal unit vector. As shown by Adler [4], eqs. (15) can be rewritten in manifestly flux conserving form by introducing a flux function $\Phi(\rho, z)$, in terms of which $\boldsymbol{D}$ is given by

$$\boldsymbol{D} = -(1/2\pi)\nabla\theta\times\nabla\Phi = -(\hat{\boldsymbol{\theta}}/2\pi\rho)\times\nabla\Phi$$

$$= \nabla\times[(\hat{\boldsymbol{\theta}}/2\pi\rho)\Phi]. \tag{16}$$

The physical interpretation of Φ follows from calculating the total flux through a surface of revolution S (with element of area $\mathrm{d}\boldsymbol{A}$) bounded by a circle C of radius ρ and z-intercept z (with element of arc-length $\mathrm{d}\boldsymbol{l} = \mathrm{d}l\hat{\boldsymbol{\theta}}$),

‡2 The analysis of refs. [3,4] used a euclidean version of the variational principle of eq. (4), in which $\mathcal{F} = (\boldsymbol{E}^a)^2 - (\boldsymbol{B}^a)^2$ was replaced by $(\boldsymbol{E}^a)^2 + (\boldsymbol{B}^a)^2$. The two descriptions coincide in the interior of the confinement domain, where $\boldsymbol{B}^a = 0$.

$$\text{flux through S} = \int_S \mathrm{d}\boldsymbol{A}\cdot\boldsymbol{D}$$

$$= \int_S \mathrm{d}\boldsymbol{A}\cdot\boldsymbol{\nabla}\times[(\hat{\boldsymbol{\theta}}/2\pi\rho)\Phi] = \int_C \mathrm{d}\boldsymbol{l}\cdot(\hat{\boldsymbol{\theta}}/2\pi\rho)\Phi = \Phi. \tag{17}$$

From eq. (17) we learn that Φ assumes the following boundary values on the axis of rotation and at infinity [‡3],

$$\Phi = 0, \quad \rho = 0, \quad |z| > a,$$

$$\Phi = Q \quad \rho = 0, \quad |z| < a,$$

$$\Phi \to 0 \quad \text{as } \rho^2 + z^2 \to \infty, \tag{18}$$

which together with eq. (16) guarantee that the equation $\boldsymbol{\nabla}\cdot\boldsymbol{D} = j_0$ is satisfied. To get a differential equation for Φ we rewrite the equation $\boldsymbol{\nabla}\times\boldsymbol{E} = 0$ in the form

$$\boldsymbol{\nabla}\cdot\{(\hat{\boldsymbol{\theta}}/\rho)\times[(-\hat{\boldsymbol{\theta}}/2\pi\rho\epsilon[D])\times\boldsymbol{\nabla}\Phi]\} = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\theta\times\boldsymbol{E})$$

$$= \boldsymbol{E}\cdot(\boldsymbol{\nabla}\times\boldsymbol{\nabla}\theta) - \boldsymbol{\nabla}\theta\cdot(\boldsymbol{\nabla}\times\boldsymbol{E}) = 0. \tag{19}$$

Expanding out the triple product on the left-hand side of eq. (19) and using $\hat{\boldsymbol{\theta}}\cdot\boldsymbol{\nabla}\Phi = 0$, we obtain

$$\boldsymbol{\nabla}\cdot[\sigma(\rho, |\boldsymbol{\nabla}\Phi|)\boldsymbol{\nabla}\Phi] = 0, \tag{20a}$$

$$\sigma(\rho, |\boldsymbol{\nabla}\Phi|) \equiv 1/\rho^2\epsilon[D], \tag{20b}$$

$$D = |\boldsymbol{\nabla}\Phi|/2\pi\rho, \quad \epsilon[D] \equiv \epsilon(E(D)),$$

with $E(D)$ obtained by inverting the equation

$$D = E\epsilon(E) = \partial\mathcal{L}_{\text{eff}}(E^2)/\partial E. \tag{21}$$

Eqs. (18)–(21) are the basic statement of the boundary value problem for Φ. Once Φ has been determined, the static potential can be calculated by substituting $\boldsymbol{\nabla}\cdot\boldsymbol{D} = j_0$ into eq. (9) and integrating by parts, giving

$$V_{\text{static}}$$

$$= \int \mathrm{d}^3x\,[ED - \mathcal{L}_{\text{eff}}(E(D)^2) + \mathcal{L}_{\text{eff}}(E(0)^2)], \tag{22}$$

where an infinite constant has been added to eq. (22) to render the integral convergent at infinity. By using the identity

$$\mathcal{L}_{\text{eff}}(E(D)^2) - \mathcal{L}_{\text{eff}}(E(0)^2)$$

$$= \int_{E(0)}^{E(D)} \mathrm{d}E'\,\partial\mathcal{L}_{\text{eff}}/\partial E' = \int_{E(0)}^{E(D)} \mathrm{d}E'\,D(E')$$

$$= ED - \int_0^D E(D')\,\mathrm{d}D', \tag{23}$$

eq. (22) can also be rewritten in the familiar form

$$V_{\text{static}} = \int \mathrm{d}^3x\,\mathcal{E}(x), \quad \mathcal{E}(x) = \int_0^D \mathrm{d}D'\,E(D'). \tag{24}$$

Considerable insight into the behavior of the solutions of eq. (20) is obtained by rewriting it to explicitly show the structure of the second derivative terms (including those arising from the $|\boldsymbol{\nabla}\Phi|$-dependence of σ). Defining the inward-directed unit normal $\hat{\boldsymbol{n}}$ and the corresponding normal derivative ∂_n,

$$\hat{\boldsymbol{n}} = \boldsymbol{\nabla}\Phi/|\boldsymbol{\nabla}\Phi|, \quad \partial_n = \hat{\boldsymbol{n}}\cdot\boldsymbol{\nabla}, \tag{25}$$

a straightforward calculation shows that eq. (20) is equivalent to

$$[\partial_\rho^2 + \partial_z^2 + (\alpha - 1)\partial_n^2]\Phi - \alpha\rho^{-1}\partial_\rho\Phi = 0, \tag{26a}$$

$$\alpha = \alpha(\rho, |\boldsymbol{\nabla}\Phi|) = 1 + \partial\log\sigma/\partial\log|\boldsymbol{\nabla}\Phi|. \tag{26b}$$

Letting $\hat{l}$ be the unit tangent to the surfaces of constant Φ and ∂_l be the corresponding tangential derivative, we have

$$\partial_\rho^2 + \partial_z^2 = \partial_n^2 + \partial_l^2 + \text{first derivative terms}, \tag{27}$$

and so the characteristic form of eq. (26a) can be written as

$$\partial_l^2 + \alpha\partial_n^2. \tag{28}$$

Thus, eq. (26a) is elliptic, parabolic or hyperbolic according as to whether $\alpha > 0$, $\alpha = 0$ or $\alpha < 0$. Before proceeding further with the analysis of the characteristics, let us use eq. (26a) to study the structure of z-axis translation-invariant solutions, for which $\Phi = \Phi(\rho)$. We then have $\partial_z\Phi = 0$, $\partial_n^2\Phi = \partial_\rho^2\Phi$, and eq. (26a) reduces to

[‡3] These boundary values correspond to the convention of always drawing the surface S so that it crosses the z-axis at a point $z_0 > a$.

$$\alpha(\partial_\rho^2 - \rho^{-1}\partial_\rho)\Phi = 0, \tag{29}$$

which on branch (I) (where $\alpha \not\equiv 0$) has the solution

$$\Phi = \Phi_0 + \pi D_0 \rho^2. \tag{30}$$

Eq. (30) describes a uniform flux $D = \hat{z} D_0$ filling all of space; we see that eqs. (20) and (26) do not admit solutions describing a translation-invariant bounded flux tube.

The analysis of eqs. (8), (9), (11) and (13)–(30) applies generally to any effective action density of the form $\mathcal{L}_{\text{eff}}(E^2)$; let us now specialize to the case of the leading logarithm model, as given in eqs. (10) and (12). Eqs. (20b), (21) and (26b), which implicitly define σ and α, are conveniently written in terms of a dimensionless function $f(w)$ defined as the solution to the transcendental equation [‡4]

$$w = f \log f, \quad f \geqslant 1, \tag{31}$$

giving

$$\sigma(\rho, |\nabla\Phi|) = (2\pi\kappa/\rho|\nabla\Phi|) f(w),$$

$$\alpha = wf'(w)/f(w), \quad w = \partial_n \Phi/\pi b_0 \kappa\rho = 2D/b_0\kappa. \tag{32}$$

For small w and large w, the behavior of $f(w)$ is given by

$$f = 1 + w + O(w^2), \quad |w| \ll 1,$$

$$f = (w/\log w)[1 + O(\log\log w/\log w)], \quad w \gg 1, \tag{33}$$

giving for the corresponding behavior of α

$$\alpha = w + O(w^2), \quad |w| \ll 1,$$

$$\alpha = 1 - (\log w)^{-1} + O(\log\log w/(\log w)^2), \quad w \gg 1. \tag{34}$$

Comparing eqs. (34), (32) and (28), we see that the differential equation of eq. (26a) is of degenerating elliptic type [5], and has a real characteristic at a surface of constant Φ where $D \propto |\nabla\Phi| = 0$. The second normal derivative $\partial_n^2\Phi$ is discontinuous across this characteristic, which acts as a free boundary, dividing space into two causally disconnected regions. From the boundary condition of eq. (18) and the continuity of Φ, we learn that $\Phi = 0$ on the free boundary. Since the exterior of the free boundary is completely surrounded by surfaces on which $\Phi = 0$, Φ vanishes identically outside the characteristic, and therefore the exterior solution lies on branch (II) of eq. (14). At a point B on the free boundary where $\rho = \rho_B$ and where the radius of curvature of the free boundary is R_B, eq. (26a) can be integrated to give an approximation to the interior solution,

$$\Phi \approx \tfrac{1}{2}(\pi b_0 \kappa \rho_B/R_B)(n - \tfrac{1}{2} l^2/R_B)^2, \tag{35}$$

with n and l normal and tangential cartesian coordinates which are zero at B. Since Φ is increasing towards the interior, we must have $R_B > 0$, and so the free boundary is everywhere convex. This in turn implies that all points on the free boundary lie within a finite distance of the origin, with the free boundary intersecting the z axis at points $\rho = 0$, $z = \pm z_B$. Since $\phi = 0$ at $z = 0$ and, from eq. (12), $E = |\nabla\phi| = \kappa$ on the free boundary, we have $\phi(\rho = 0, z = \pm z_B) = \pm\kappa L$, with L the length of the segment of the free boundary lying in the quadrant $\rho > 0$, $z > 0$ of the ρ, z plane. But since the scalar potential ϕ becomes infinite at the source charge coordinates $\rho = 0$, $z = \pm a$, this implies that $z_B > a$. The only qualitative feature of the solution which we have been unable to characterize analytically is the detailed structure of the free boundary–z-axis intersection; the numerical results strongly suggest that the free boundary is smooth, with no cusp at the z-axis, but we do not have a proof of this. Turning finally to the static potential, we learn from eqs. (31) and (32) that inside the free boundary E is bounded by

$$E(D)/\kappa = f(w) \geqslant 1, \tag{36}$$

which when substituted into eq. (24) gives

$$V_{\text{static}} \geqslant \kappa \int d^3x\, D. \tag{37}$$

Writing $d^3x = dl\, dA$, with l the length along and dA the element of area perpendicular to the flux lines of D, eq. (37) yields the lower bound [‡5]

‡4 The choice of branch is again dictated by the requirement that the solution be continuously connected to the strong-field, asymptotically free regime.

‡5 For a discussion of the removal of the infinite Coulomb self-energies from eq. (38), see ref. [3].

Volume 113B, number 5 PHYSICS LETTERS 1 July 1982

$$V_{\text{static}}(R) \geqslant \kappa \int \mathrm{d}A\, D \int \mathrm{d}l \geqslant \kappa Q\, l_{\min} = \kappa Q R,$$

$$R = |x_1 - x_2| = 2a. \tag{38}$$

In determining V_{static} computationally, we use eq. (23), which in the leading logarithm model can be conveniently rewritten in the form

$$V_{\text{static}} = \int \mathrm{d}^3x\, \tfrac{1}{2}\sigma(\nabla\Phi/2\pi)^2(1+\xi),$$

$$\xi = (f^2-1)/(2fw). \tag{39}$$

We believe that the connection found above between degenerating elliptic operators and a linearly increasing static potential is a very general one. For example, let us briefly consider the case in which $\mathcal{L}_{\text{eff}}(E^2)$, rather than attaining its minimum at a non-zero value of E as in the leading logarithm model, attains its minimum at $E = 0$ and behaves there as $\mathcal{L}_{\text{eff}} \sim E^\gamma$. Then for an isolated charge at $r = 0$ in spherical coordinates, one has $r^{-2} \sim D \sim E^{\gamma-1}$, giving [2] for the infrared energy in a box of side L (and presumably for the long-distance behavior of the static potential of two opposite charges)

$$E_{\text{infrared}} \sim L^3 DE \sim L^{(\gamma-3)/(\gamma-1)}. \tag{40}$$

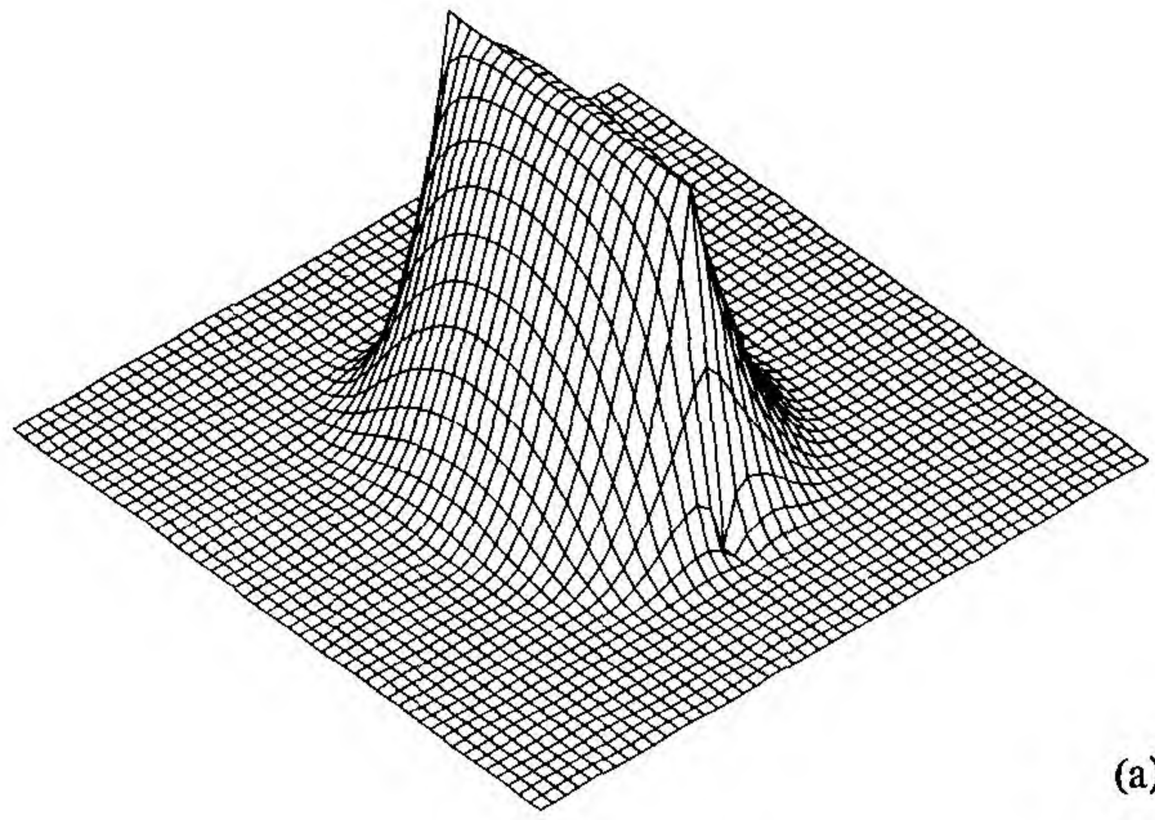
Substituting $\mathcal{L}_{\text{eff}} \sim E^\gamma$ into eqs. (21)–(26) above, we find that

$$\alpha(D=0) = 1 + \partial(\log E^{2-\gamma})/\partial(\log E^{\gamma-1}) = (\gamma-1)^{-1}. \tag{41}$$

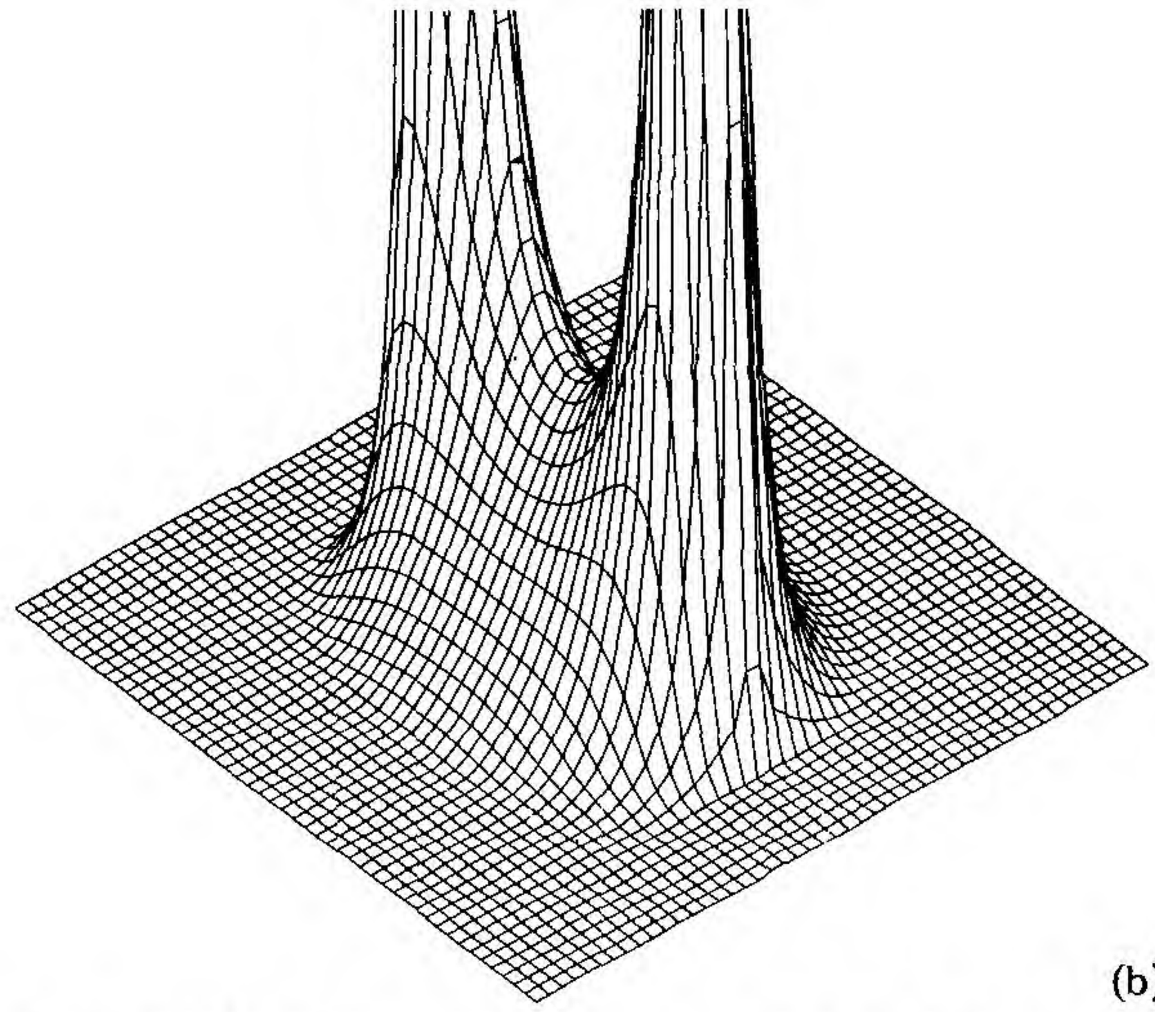
Thus for finite γ, where the infrared energy grows less strongly than linearly with L, the characteristic form of eq. (26a) is always elliptic, while in the limit as $\gamma \to \infty$, where the infrared energy grows linearly, the characteristic form again degenerates [‡6] from elliptic to parabolic at $D = 0$.

3. Numerical results. We have developed a numerical method for solving eq. (20), based on a two-step procedure in which Φ is updated by a single over-relaxed iteration of eq. (20a) for fixed σ, and then σ is updated by using eqs. (20b), (31), and (32), with a Newton iteration used to solve the transcendental equation for f. The boundary conditions of eq. (18) are applied on the rotation axis, while the boundary condition $\Phi = 0$ is imposed on the perimeter of the computational grid, which must be chosen large enough to completely enclose the ultimate free boundary. A detailed discussion of the theoretical and practical aspects of the numerical algorithm will be given elsewhere [6]; we give here some sample results from our calculations, done for $Q = (4/3)^{1/2}$ and $b_0 = 9/8\pi^2$

‡6 These estimates do not determine the structure of the confinement domain. Very likely, when $\mathcal{L}_{\text{eff}}$ has an infinite order zero at $E = 0$, the confinement domain fills all of space. (We wish to thank J. Chayes and L. Chayes for discussions about this case.) For a renormalization group argument suggesting that the minimum of $\mathcal{L}_{\text{eff}}$ in fact lies at non-zero E, see ref. [4].



(a)



(b)

Fig. 1. (a) The flux function Φ on a plane through the z-axis, plotted vertically on a linear scale. The base of the figure is at $\Phi = 0$. (b) The energy density $\mathcal{E}$, similarly plotted.

Volume 113B, number 5 PHYSICS LETTERS 1 July 1982

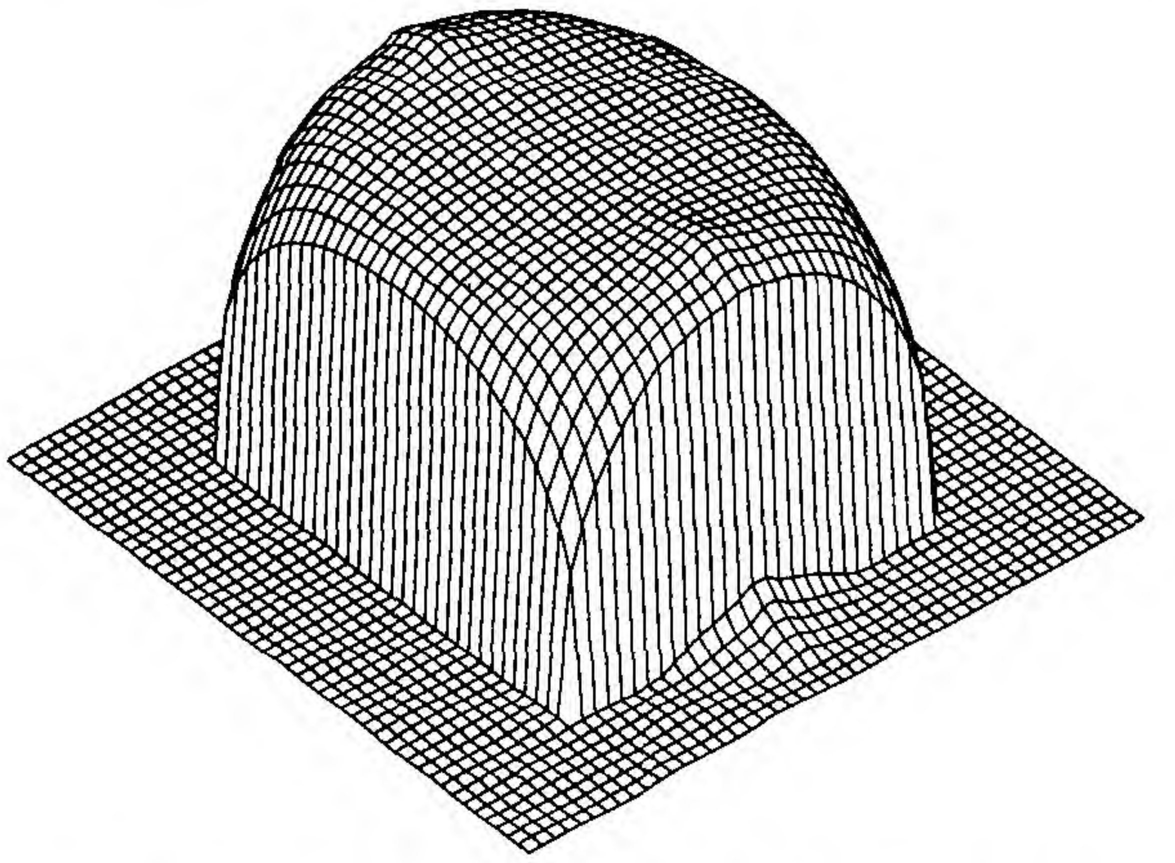


Fig. 2. The dielectric constant ϵ on a plane through the z-axis, plotted vertically on a logarithmic scale. From the top of the figure to the base spans 14 decades, corresponding to $\epsilon_{\min} = 10^{-15}$. (When $\epsilon_{\min}$ is reduced to 10^{-35}, the residual structure along the axis at the base of the figure is eliminated.)

[corresponding to SU(3) QCD with three light quark flavors]. The results shown in figs. 1 and 2 were computed for $\kappa^{1/2}R = 8$, and required roughly 1–2 min of CPU time on a VAX 11/780 computer, for convergence on a single-quadrant 25 × 25 computational mesh. The figures all show elevation plots of physical quantities measured on a plane passing through the rotation axis. Fig. 1a shows the flux function Φ plotted on a linear scale starting at $\Phi = 0$; the discontinuity of eq. (18) was enforced by taking the charges on lattice sites where $\Phi = Q/2$, and taking $\Phi = Q$ ($\Phi = 0$) on the axis for $|z| < a$ ($|z| > a$), as is clearly visible on the plot. Fig. 1b shows the field energy density $\mathcal{E}(x)$ plotted on a linear scale starting at $\mathcal{E} = 0$, with the Coulomb energy peaks (which rise by a further factor of 100 before being cut off by the finite mesh size) clearly visible. From fig. 1 we see that Φ and $\mathcal{E}$ are nonzero only within an oval-shaped curve, which is the continuum-limit free boundary, and which crosses the axis of rotation with no visible cusp in $\mathcal{E}$. The complete independence of the interior and exterior regions can be seen from fig. 2, which shows ϵ plotted on a logarithmic scale ranging from an imposed lower limit of $\epsilon_{\min} = 10^{-15}$ at the base to a maximum (governed by the mesh spacing) of $\epsilon \sim 0.1$ at the Coulomb peaks. A logarithmic plot of $\mathcal{E}$ looks very similar, apart from having more pronounced dimples at the charge sites. Repeating the computation on meshes up to a factor of 4 finer shows good convergence as the mesh spacing goes to zero [‡7]. We find that in the range $0.25 \leqslant \kappa^{1/2} R \leqslant 128$, the large-distance behavior of $V_{\mathrm{static}}(R)$ is fitted by the formula

$$V_{\mathrm{static}}(R) = (4/3)^{1/2}\kappa R + 1.95\,\kappa^{1/2}\log(\kappa^{1/2}R)$$

$$+ \text{constant} - 0.40/R, \tag{42}$$

with uncertainties of order 1 in each final decimal place. Thus, the computational results indicate that the bound of eq. (38) is saturated and that the leading correction to the linear potential is logarithmic, in marked contrast to the $AR + B + CR^{-1}$ form expected [7] in the string model and in strong-coupling lattice gauge theories. This feature of the effective action approach should eventually have testable experimental consequences. In subsequent work we plan to map out V_{static} for all distances and to determine the ratio of the string tension to $\Lambda_{\overline{\mathrm{MS}}}$, in both the leading logarithm model and in the extended model in which log log renormalization group corrections are included in ϵ.

This work was supported by the Department of Energy under Grant Number DE-AC02-76ER02220.

‡7 More precisely, to eliminate the Coulomb self-energy divergences we study only *differences* $V_{\mathrm{static}}(R_1) - V_{\mathrm{static}}(R_2)$, with identical mesh structure around the charges. These quantities converge as the mesh spacing is decreased.

References

[1] S.L. Adler, Phys. Lett. 110B (1982) 302.
[2] H. Pagels and E. Tomboulis, J. Phys. B143 (1978) 485.
[3] S.L. Adler, Phys. Rev. D23 (1981) 2905.
[4] S.L. Adler, The mechanism for confinement in massive quark QCD, in: Proc. 5th Workshop on current problems in particle theory (Johns Hopkins University, Baltimore, 1981).
[5] O.A. Oleĭnik and E.V. Radkevič, Second order equations with nonnegative characteristic form (American Mathematical Society, Providence – Plenum, New York, 1973).
[6] S.L. Adler and T. Piran, Rev. Mod. Phys., in preparation.
[7] M. Lüscher, Nucl. Phys. B180 (1981) 317.

ERRATA

S.L. Adler and T. Piran, Flux confinement in the leading logarithm model, Phys. Lett. 113B (1982) 405.

On page 406, first line below eq. (6), "confinement" should read "coefficient".

On page 40, ref. [2], "J. Phys." should read "Nucl. Phys.".

Ref. [6] should read: S.L. Adler and T. Piran, Relaxation methods for gauge field equilibrium equations, Rev. Mod. Phys., to be published.

Volume 117B, number 1, 2 PHYSICS LETTERS 4 November 1982

THE HEAVY QUARK STATIC POTENTIAL IN THE LEADING LOG AND THE LEADING LOG LOG MODELS

Stephen L. ADLER and Tsvi PIRAN [1]
The Institute for Advanced Study, Princeton, NJ 08540, USA

Received 21 June 1982

We give numerical and analytical results for the static potential of quasi-abelian quark and antiquark source charges, in the models in which radiative corrections to leading log order, and to leading log plus leading log log order, are retained in the gauge gluon effective dielectric functional. When the scale length of the latter model is fixed to give a second order fit to Martin's phenomenological heavy quark potential, the point of tangency with Martin's curve lies at 0.51 fermi (in the center of the $c\bar{c}$ and $b\bar{b}$ quarkonium region), and the long- and short-distance limits yield, respectively, a string tension of 320 MeV, and an asymptotic freedom scale mass of $\Lambda_{\overline{\mathrm{MS}}} = 220$ MeV [all for color SU(3) with 3 light quark flavors]. Thus, effective action methods, using only renormalization group-improved perturbation theory as input, give a reasonable account of the behavior of the heavy quark–antiquark static potential at all length scales.

In a recent letter [1], we gave a qualitative analysis of non-linear effective action models for heavy quark statics, and showed that in the renormalization group approximation they predict both total flux confinement, and a linear static potential at large source separations. As a continuation of our study of the renormalization group models, we present here results of a numerical computation of the static potential at all length scales, from which we extract theoretical predictions for several parameters characterizing the strong interactions.

The models under consideration can be written as a problem in non-linear electrostatics,

$$\nabla \cdot D = j_0 , \quad \nabla \times E = 0 , \tag{1}$$

with j_0 a source density corresponding to static, quasi-abelian quark and antiquark source charges,

$$j_0 = Q[\delta^3(x - x_1) - \delta^3(x - x_2)] , \quad Q = (4/3)^{1/2} , \quad |x_1 - x_2| = R , \tag{2}$$

and with D related to E by the non-linear constitutive equation

$$D = \epsilon(E)E . \tag{3}$$

In the leading log and leading log log models, $\epsilon(E)$ is given by [‡1]

$$\epsilon(E) = \tfrac{1}{2} b_0 \log(E/\kappa), \quad \text{leading log} , \tag{4a}$$

$$\epsilon(E) = \tfrac{1}{2} b_0 [\log(E/\kappa) + \xi \log\log(E/\kappa)] , \text{leading log log}, \tag{4b}$$

with b_0 and ξ the renormalization group constants [for SU(3) quantum chromodynamics (QCD) with N_f light quark flavors]

$$b_0 = (1/8\pi^2)(11 - \tfrac{2}{3}N_f) , \quad \xi = 2(51 - \tfrac{19}{3}N_f)/(11 - \tfrac{2}{3}N_f)^2 . \tag{5}$$

Throughout our calculations we will take $N_f = 3$, giving

[1] Also at the Racah Institute of Physics, The Hebrew University, Jerusalem, Israel.

‡1 The effective action corresponding to eq. (4a) contains a single leading logarithm term. The effective action corresponding to eq. (4b) contains the standard renormalization group log and log log terms, plus additional subdominant terms which vanish as $|\log(E/\kappa)|$ becomes infinite. Alternatively, if the log log model were defined by taking the effective action functional to have exactly a log plus log log form, the corresponding effective dielectric functional would differ from eq. (4b) by subdominant terms.

Volume 117B, number 1, 2 PHYSICS LETTERS 4 November 1982

$$b_0 = 9/8\pi^2 = 0.1140, \quad \xi = 0.790\,. \tag{6}$$

This fixes all parameters in the model apart from the scale mass $\kappa^{1/2}$, which is set equal to unity to define dimensionless units for the numerical computations, and then is reinserted below and determined by comparing with Martin's fit to the heavy quarkonium spectra. According to the standard renormalization group analysis [2], eq. (4b) gives the leading two terms in an asymptotic expansion of $\epsilon(E)$ for strong fields E. At strong fields, the corrections to the expression in square brackets in eq. (4b) are expected to be of order unity and, to the extent that they are slowly varying, can be absorbed into the scale mass $\kappa^{1/2}$. We will, however, be using the formulas of eqs. (4a) and (4b) outside the strong field region, in fact for all fields $E_{\min} \leqslant E < \infty$, where $\epsilon(E_{\min}) = 0$. A rationale for this extension has been given by Adler [3], who points out that renormalization group estimates are formally valid whenever the running coupling

$$g^2(E) \sim [\tfrac{1}{2} b_0 \log(E/\kappa) + \ldots]^{-1}\,,$$

is small in magnitude, which is true both when $E/\kappa \gg 1$ [giving $g^2(E)$ small and positive] and when $E/\kappa \ll 1$ [giving $g^2(E)$ small and negative]. This argument suggests that there is a second *weak field* asymptotically free regime, where eq. (4b) again gives the leading behavior of ϵ, and where ϵ (or more precisely, the real part of ϵ) is negative. Although renormalization group estimates cannot be used at intermediate field strengths, the statements that $\epsilon < 0$ for $E/\kappa \ll 1$ and $\epsilon > 1$ at $E/\kappa \gg 1$ imply that ϵ must cross from 1 to 0 at some intermediate field strength $E_{\min}$, and this is the essential feature of our model. The numerical results obtained below suggest that, in fact, the corrections to eq. (4b) are of the form

$$\epsilon(E) = \tfrac{1}{2} b_0 \,[\log(E/\kappa) + \xi \log\log(E/\kappa) + \mathrm{O}(1)]\,, \tag{7}$$

with O(1) representing non-constant terms which are effectively of order unity in the entire range $E_{\min} \leqslant E < \infty$ ‡2.

As discussed in ref. [1], the numerical procedure for solving the model of eqs. (1)–(4) treats the D field as the fundamental dynamical variable, in terms of which E is obtained by inverting the constitutive equations of eq. (4). This gives

‡2 For footnote see next column.

$$E/\kappa = f(w)\,, \quad w = 2D/b_0\kappa\,, \tag{8a}$$

with $f(w)$ a transcendental function defined by

$$\begin{aligned} w &= f \log f\,, && \text{log model}\,, \\ w &= f(\log f + \xi \log\log f)\,, && \text{log log model}\,, \end{aligned} \tag{8b}$$

from which we can calculate the minimum electric field

$$\begin{aligned} E_{\min}/\kappa = f(0) &= 1 && \text{log model}\,, \\ &= 1.680 && \text{log log model}\,. \end{aligned} \tag{8c}$$

Once the D and E fields have been solved for by the numerical algorithm of ref. [1], the static potential can be calculated from the formula

$$\begin{aligned} V_{\text{static}}(R) &= \int \mathrm{d}^3x \int_0^D E(D')\,\mathrm{d}D' \\ &= \int \mathrm{d}^3x\, \tfrac{1}{2} b_0 \kappa^2 \{\tfrac{1}{2} w f(w) + \tfrac{1}{4}[f(w)^2 - f(0)^2] \\ &\quad + \tfrac{1}{2}\xi[y(2\log f(w)) - y(2\log f(0))]\}\,, \end{aligned} \tag{9}$$

with $y(x)$ the exponential integral

$$y(x) = \int_{-\infty}^{x} \frac{\mathrm{e}^t}{t}\,\mathrm{d}t\,, \tag{10}$$

which is available in the IMSL function library. Some useful results concerning the large-R and small-R limits of V_{static} can be deduced by analytical methods. A simple flux conservation argument [1,6] shows that the static potential has a linear lower bound for large R,

‡2 Possible theoretical support for the idea of a weak field asymptotically free regime is given by recent work of 't Hooft [4] and of De Calan and Rivasseau [5], showing that the sum of planar skeleton diagrams has a finite radius of convergence as a function of coupling constant. Assuming that renormalization effects can be taken into account by using running couplings in the skeleton expansion, this result implies that in the large-N_c limit, where planar diagrams dominate, the color dielectric constant will be an analytic function of the running coupling for small $|g^2(E)|$. Renormalization group estimates will then apply for small negative as well as small positive $g^2(E)$, and if the $N_c = 3$ theory behaves qualitatively like the $N_c = \infty$ limit, one is led to the model discussed in the text.

$$V_{\rm static}(R) \geqslant E_{\rm min} QR + \text{const}\,, \quad R \to \infty, \tag{11}$$

and our numerical results (accurate to a few tenths of a percent, and extending out to $\kappa^{1/2}R$ or order 100) show that this bound is saturated. Hence we have

$$\text{string tension} = \kappa^{1/2}\,[Qf(0)]^{1/2}\,. \tag{12}$$

A detailed analysis [7] of the short-distance perturbation theory of the leading log and log log models shows that

$$V_{\rm static}(R) = -(Q^2/4\pi R\,\tfrac{1}{2}b_0)\,[\zeta + O(\zeta^3)]\,, \quad R \to 0,$$

$$\zeta = f(w_{\rm P})/w_{\rm P}, \qquad w_{\rm P} \equiv 1/\Lambda_{\rm P}^2 R^2\,, \tag{13}$$

with $\Lambda_{\rm P}$ given by a numerical integral, the evaluation of which [7] yields

$$\Lambda_{\rm P} = 2.52\,\kappa^{1/2}\,. \tag{14}$$

Since from eq. (8b) we find

$$\zeta = 1/\log w_{\rm P} + O[\log\log w_{\rm P}/(\log w_{\rm P})^2] + O[(1/\log w_{\rm P})^3]\,, \tag{15}$$

$\Lambda_{\rm P}$ corresponds to the standard definition of the scale mass associated with the static potential ‡3, which is in turn related to the scale mass $\Lambda_{\rm r}$ associated with the force law,

$$-V'_{\rm static}(R) = (Q^2/4\pi R^2\,\tfrac{1}{2}b_0)\{1/\log w_{\rm F} + O[\log\log w_{\rm F}/(\log w_{\rm F})^2] + O[1/(\log w_{\rm F})^3]\}\,,$$

$$w_{\rm F} \equiv 1/\Lambda_{\rm r}^2 R^2\,, \tag{16}$$

by [8]

$$\Lambda_{\rm r} = \Lambda_{\rm P}/e\,. \tag{17}$$

Since $\Lambda_{\rm r}$ is related to the standard strong interaction scale mass $\Lambda_{\overline{\rm MS}}$ by the formula [8,9] (with $\gamma_{\rm E}$ = 0.5772... Euler's constant)

$$\Lambda_{\rm r}/\Lambda_{\overline{\rm MS}} = \exp[\gamma_{\rm E} - 1 + (1/8\pi^2 b_0)(\tfrac{31}{6} - \tfrac{5}{9}N_{\rm f})]$$
$$= 0.967 \quad \text{for} \quad N_{\rm f} = 3\,, \tag{18}$$

‡3 As discussed in ref. [7], the log log model does not give the correct value of the coefficient of the log log term in the expansion of eq. (16) (where ξ should appear, the log log model gives $\xi - 1$), and so it is not good beyond one-loop order at short distances. Nonetheless, the model permits a meaningful determination of the scale mass $\Lambda_{\rm P}$, because this involves only a one-loop calculation which is independent of the value of the parameter ξ.

combining eqs. (14), (17) and (18) yields

$$\Lambda_{\overline{\rm MS}} = 0.959\,\kappa^{1/2}\,. \tag{19}$$

Taking the ratio of eq. (19) to eq. (12), the scale mass $\kappa^{1/2}$ drops out, and we get

$$\Lambda_{\overline{\rm MS}}/\text{string tension} = 0.959/[Qf(0)]^{1/2}$$
$$= 0.892 \quad \text{log model}\,,$$
$$= 0.689 \quad \text{log log model}\,. \tag{20}$$

Experimentally, if the string tension is identified with that computed from the slopes of Regge trajectories, we have

$$\text{string tension} \sim 400\ \text{MeV}\,. \tag{21a}$$

Recent determinations of $\Lambda_{\overline{\rm MS}}$ give values

$$100\ \text{MeV} \lesssim \Lambda_{\overline{\rm MS}} \lesssim 400\ \text{MeV}\,, \tag{21b}$$

$$\to 0.25 \lesssim (\Lambda_{\overline{\rm MS}}/\text{string tension})_{\rm expt} \lesssim 1\,, \tag{21c}$$

with the values at the lower end of the indicated ranges currently favored [10].

Since $V_{\rm static}(R)$ as given by eq. (9) contains infinite additive self-energy contributions, the quantity which can be measured computationally is $V_{\rm static}(R_1) - V_{\rm static}(R_2)$, where identical mesh structures around the source charges must be used at the two separations R_1, R_2. By using standard over-relaxation methods [11] we have made a sequence of measurements of $\Delta_2 V_{\rm static} \equiv V_{\rm static}(R) - V_{\rm static}(R/2)$, with $\kappa^{1/2}R$ ranging from ~ 100 to 10^{-7}, for both the leading log and the leading log log models. For $\kappa^{1/2}R$ smaller than 1 the use of jacobian transformations is necessary to get good accuracy; this and other details of the computational methods will be described in a pedagogical review article which is in preparation [12]. As already noted, the large-distance results show that the bound of eq. (11) is saturated. For the smallest R values studied ($10^{-7} < \kappa^{1/2}R < 10^{-5}$), the measurements of $\Delta_2 V_{\rm static}$ agree with the leading term of the asymptotic formula of eqs. (13) and (14) to within a few tenths of a percent. The entire sequence of $\Delta_2 V_{\rm static}$ values has been fitted to the analytical formula

$$V_{\rm static} = \kappa^{1/2}\,[F(\kappa^{1/2}R) + C]\,, \tag{22}$$

with $F_{\log}(r)$ and $F_{\log\log}(r)$ as given in table 1, and with C an undetermined overall constant. We estimate the combined accuracy of the measurements and the

Volume 117B, number 1, 2 PHYSICS LETTERS 4 November 1982

Table 1
Functions $F_{\log}(r)$ and $F_{\log\log}(r)$. The coefficients are given in table 2.

r range	$F(r)$
$r \leqslant 0.0125$	$F(r) = -(Q^2/4\pi r_2^1 b_0)\,[f(w_P)/w_P](1+a_1 r^{a_2})$ $w_P = 1/(2.52r)^2$, $f(w)$ as in eq. (8b)
$0.0125 \leqslant r \leqslant 0.125$	$F(r) = K + \alpha(r/0.125)^E$ $E = \beta + \gamma\ln(1/r) + \delta[\ln(1/r)]^2 + \epsilon[\ln(1/r)]^3$
$0.125 \leqslant r \leqslant 2$	$F(r) = K' + \alpha'\log r + \beta'(\log r)^2 + \gamma'(\log r)^3 + \delta'(\log r)^4 + \epsilon'(\log r)^5$
$2 \leqslant r$	$F(r) = K'' + (4/3)^{1/2} f(0) r + \alpha''/r^{1/2} + \beta''/r + \gamma''\log r$

fitting for $\Delta_2 V_{\text{static}}$ to be better than 2% for all R, and better than 1% for very large and very small R values.

To determine the scale mass $\kappa^{1/2}$ appearing in eq. (22), we fit the potentials of table 1 to the phenomenological formula given by Martin [13]

$$V_{\text{static}} = -8.064\ \text{GeV} + 6.870\ (\text{GeV})^{1.1} R^{0.1}\ , \qquad (23)$$

which is known to give a good description of the $c\bar{c}$ and $b\bar{b}$ quarkonium spectra. Writing $R = r/\kappa^{1/2}$, with

Table 2
Coefficients of functions $F_{\log}(r)$ and $F_{\log\log}(r)$ as given in table 1.

coefficient	log model	log log model
a_1	5.38	5.14
a_2	0.545	0.556
K	−2.323	2.569
α	−15.118	−15.717
β	−0.305	−0.305
γ	0.00368	0.06964
δ	−0.00885	−0.0284
ϵ	0.00043	0.00212
K'	−9.686	−5.950
α'	3.518	3.840
β'	0.355	0.807
γ'	0.256	0.395
δ'	0.0439	0.0683
ϵ'	0.0127	0.0110
K''	−10.520	−7.781
α''	0.139	0.465
β''	−0.46	−0.58
γ''	1.966	1.604

r dimensionless, and differentiating to eliminate the additive constant C, we find that the choice

$$\kappa^{1/2}(r^*) = [0.687/(r^*)^{0.9} F'(r^*)]^{(1/1.1)}\ \text{GeV}\ , \qquad (24)$$

makes the slope of eq. (22) identical to that of Martin's potential at $r = \kappa^{1/2}R = r^*$. A plot of eq. (24) shows that $\kappa^{1/2}(r^*)$ vanishes as $r^* \to 0$ and as $r^* \to \infty$, and has a single maximum in between. At the maximum of $\kappa^{1/2}(r^*)$, the potential of eq. (22) has a second order contact with Martin's curve, giving the closest fit. From the formulas of table 1, we find that [‡4]

$$\kappa^{1/2}_{\max} = 0.2291\ \text{GeV at } r^* = 0.789 \text{ log model,}$$
$$= 0.2274\ \text{GeV at } r^* = 0.589 \text{ log log model}. \qquad (25)$$

When reexpressed in physical distance units, the match points are

$$R^* = r^*/\kappa^{1/2}_{\max}$$
$$= (0.789/0.2291\ \text{GeV}) \times (l = 0.1973\ \text{GeV fermi})$$
$$= 0.68 \text{ fermi log model},$$

$$R^* = r^*/k^{1/2}_{\max} =$$
$$= (0.589/0.2274\ \text{GeV}) \times (l = 0.1973\ \text{GeV fermi})$$
$$= 0.51 \text{ fermi log log model}, \qquad (26)$$

which lie in the middle of the heavy quarkonium region of 0.1 fermi $< R <$ 1 fermi, thus giving a nontrivial check on both the fitting procedure and the models. Adjusting the additive constants C to bring the potentials of eq. (22) into coincidence with Martin's curve at the match points, we get the formulas

$$V_{\text{static}}(R) = 0.2291\ \text{GeV} \times [F_{\log}(0.2291\ \text{GeV}\,R) + 9.250]\ , \qquad (27a)$$

$$V_{\text{static}}(R) = 0.2274\ \text{GeV} \times [F_{\log\log}(0.2274\ \text{GeV}\,R) + 5.582]\ , \qquad (27b)$$

for the static potential in the log and log log models,

‡4 An interesting feature of the fitting procedure is that although V_{static} changes significantly from the log to the log log model, the values of the scale mass $\kappa^{1/2}$ determined in the two cases agree to better than a percent.

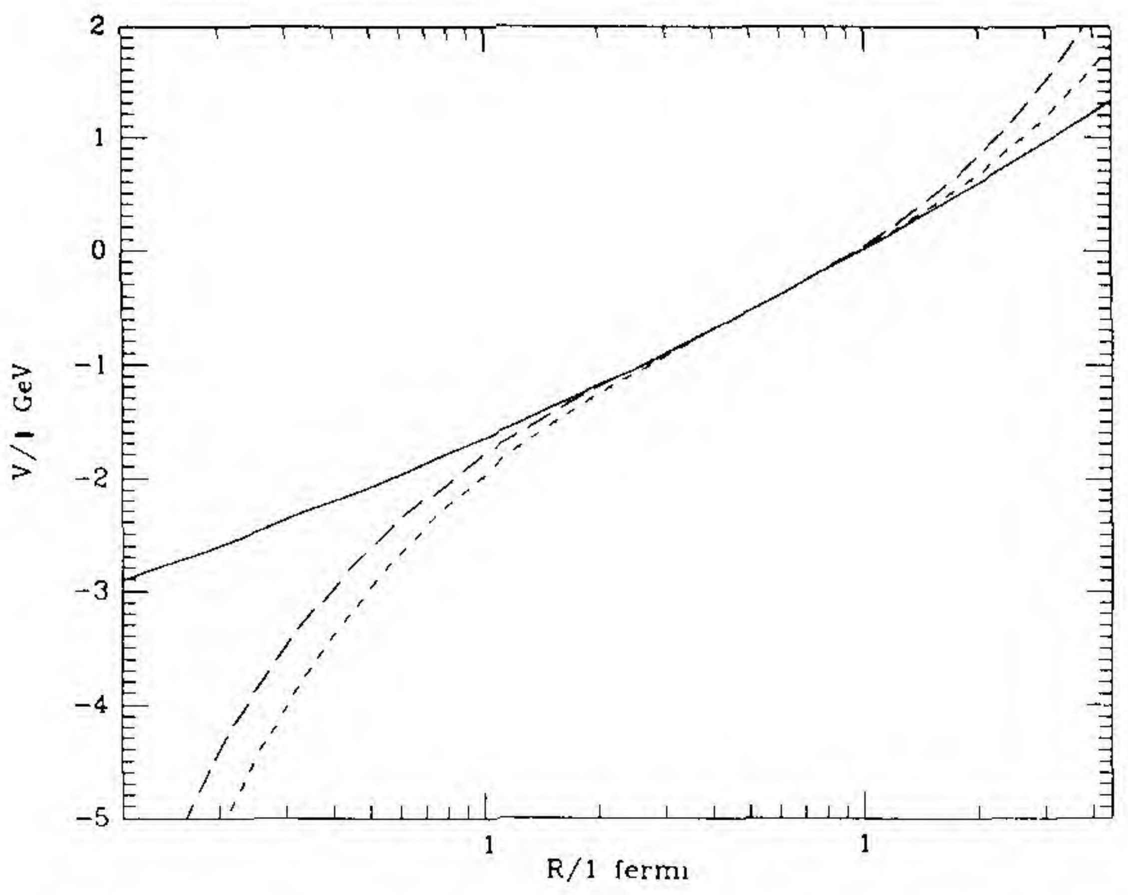


Fig. 1. The potentials of eq. (17a) [log model; short dashes], eq. (27b) [log log model; long dashes], and eq. (23) (Martin's fit; solid line). The $c\bar{c}$ and $b\bar{b}$ quarkonium region lies between 0.1 fermi and 1 fermi, and the validity of eq. (23) is restricted to this interval.

respectively. These are plotted, together with Martin's curve of eq. (23), in fig. 1. The potentials of eq. (27), particularly that for the log log model, are clearly in good agreement with Martin's curve in the heavy quarkonium region.

Substituting eq. (25) back into eqs. (12) and (19) we get the string tension and $\Lambda_{\overline{MS}}$ in physical units,

$$\text{string tension} = 250\ \text{MeV} \quad \text{log model},$$
$$= 320\ \text{MeV} \quad \text{log log model},$$
$$\Lambda_{\overline{MS}} = 220\ \text{MeV} \quad \text{both models}. \qquad (28)$$

To summarize, effective action models, which use only renormalization group-improved perturbation theory as input (but which employ nonperturbative methods to solve the resulting differential equations), give a reasonable account of the heavy quark static potential at all length scales. This suggests that the following directions for further investigation using these methods will be of interest: (i) The study of the spin–spin and spin–orbit potentials in heavy quark systems. A formalism for doing this has been set up by Hiller [14], and will give parameter-free predictions using the potentials of eq. (27) as input. (ii) The use of the relativistic effective action model proposed by Adler [15] to study binding in light-quark systems, and in particular to investigate chiral symmetry breaking effects. (iii) The study of whether a rearrangement [15] of QCD around a zeroth order approximation based on the use of the renormalization group-improved effective action can be used to give a systematic approximation scheme for treating bound state problems, which are currently inaccessible using the standard perturbative QCD methods.

We wish to thank H. Pagels for calling our attention to ref. [13]. This work was supported by the Department of Energy under Grant Number DE-AC02-76ER02220.

Note added. Where the term "string tension" is used in the text, what is meant is the *square root* $\sigma^{1/2}$ of the string tension as conventionally defined by the formula

$$V_{\text{static}}(R) \underset{R\to\infty}{=} \sigma R + O(1)\,.$$

References

[1] S.L. Adler and T. Piran, Phys. Lett. 113B (1982) 405.
[2] H. Pagels and E. Tomboulis, Nucl. Phys. B143 (1978) 485.
[3] S.L. Adler, in: Proc. 5th Johns Hopkins Workshop on Current problems in particle theory (Johns Hopkins, Baltimore, 1981).
[4] G. 't Hooft, On the convergence of planar diagrams, Utrecht preprint (1982); Phys. Lett. 109B (1982) 474.
[5] C. de Calan and V. Rivasseau, Commun. Math. Phys. 82 (1981) 69.
[6] S.L. Adler, Phys. Rev. D23 (1981) 2905.
[7] S.L. Adler, in preparation.
[8] E. Kovacs, A Monte Carlo evaluation of the interquark potential, Rockefeller University report RU82/B/18.
[9] W. Buchmüller and S.-H.H. Tye, in: Perturbative quantum chromodynamics, D.W. Duke and J.F. Owens, eds. (Tallahassee, 1981), AIP Conference Proceedings No. 74 (AIP, New York) p. 370;
A. Billoire, Phys. Lett. 104B (1981) 472.
[10] P.B. MacKenzie and G.P. Lepage, in: Perturbative quantum chromodynamics, D.W. Duke and J.F. Owens, eds. (Tallahassee, 1981), AIP Conference Proceedings, No. 74 (AIP, New York).
[11] W.F. Ames, Numerical methods for partial differential equations (Academic Press, New York, 1977).
[12] S.L. Adler and T. Pira, to be published in Rev. Mod. Phys.
[13] A. Martin, Phys. Lett. 100B (1981) 511.
[14] J.R. Hiller, An effective action approach to spin potentials for heavy quark systems, Ann. Phys. (N.Y.), to be published.
[15] S.L. Adler, Phys. Lett. 110B (1982) 302.

S.L. Adler and T. Piran, The heavy quark static potential in the leading log and the leading log log models, Phys. Lett. 117B (1982) 91.

On page 92, first column, the argument for the existence of an $E_{\min}$ at which $\epsilon(E_{\min}) = 0$ implicitly makes the physically motivated assumption that $\epsilon(E)$ is a continuous function of E. Continuity of ϵ, together with the statements $\epsilon < 0$ for $E/\kappa \ll 1$ and $\epsilon > 1$ for $E/\kappa \gg 1$, implies that ϵ vanishes somewhere in the intermediate-field-strength region where $E \sim \kappa$ and where the running coupling is large, even though renormalization group arguments cannot be directly used there. In a recent letter by Elizalde [E. Elizalde, Is the next-to-leading log model confining?, Phys. Lett. 115B (1982) 307], renormalization group arguments are used uncritically in the intermediate field strength region. This leads to a spurious infinite discontinuity in $\epsilon(E)$, and hence to Elizalde's erroneous conclusion that log log renormalization group corrections spoil the confining property of the leading log model.

On page 94, in the first line of table 1, "$^1_2$" should read "$\frac{1}{2}$".

In both parts of eq. (26), "l = 0.1973 GeV fermi" should read "1 = 0.1973 GeV fermi", and in the second part of eq. (26), "$k_{\max}^{1/2}$" should read "$\kappa_{\max}^{1/2}$".

On page 95, in the first line of the caption to fig. 1, "eq. (17a)" should read "eq. (27a)".

Ref. [7] should read: S.L. Adler, Short distance perturbation theory for the leading logarithm models, Nucl. Phys. B., to be published.

Ref. [12] should read: S.L. Adler and T. Piran, Relaxation methods for gauge field equilibrium equations., Rev. Mod. Phys., to be published.

PHYSICAL REVIEW D VOLUME 27, NUMBER 8 15 APRIL 1983

Quasi-Abelian versus large-N_c linear confinement

Stephen L. Adler and Herbert Neuberger
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 6 December 1982)

In the quasi-Abelian approximation as well as in the large-N_c limit sources in the adjoint representation cannot be screened. Nevertheless, the two phenomena are different because the string tension is proportional to the square root of the Casimir operator in the first case and to the Casimir operator itself in the second.

It is possible that the response of a non-Abelian gauge theory to the introduction of external color sources can be modeled by the response of a nonlinear Abelian system to appropriately defined external charges. If we restrict ourselves to local and gauge-invariant systems the model is defined by an action of the following type:

$$\mathfrak{A} = \int d^4x \{f[F_{\mu\nu}{}^2(x)]F_{\mu\nu}{}^2(x) + j_\mu^{\rm ext}A_\mu\} \quad . \tag{1}$$

By locality we mean that f depends only on $F_{\mu\nu}{}^2(x)$ and not on higher derivatives. Systems of the type (1) have been investigated lately from an effective-action viewpoint.[1] In a limiting case they reduce to the models used in bag computations.[2]

The Abelian model encounters one obvious difficulty: The medium cannot screen even zero-triality [for $SU(3)_c$] sources. Therefore, the existence of linear forces binding quarks implies the same for color octets separated by arbitrary large distances. The Abelian vacuum cannot simulate color-octet pair creation.

It was noted some time ago that a somewhat similar phenomenon occurs in the $N=\infty$ limit of a non-Abelian $SU(N)$ gauge theory[3]: At infinite N the probability of creating a pair of particles in the adjoint representation with the color content necessary to screen a fixed external charge vanishes.[4] Again the existence of linear forces binding charges in the fundamental representation implies the same for the adjoint representation.

The purpose of this Brief Report is to point out that the above similarity does not work on a more quantitative level.

We start by using an argument due to Lieb regarding Eq. (1).[5] For the static problem we take

$$j_\mu^{\rm ext} = \delta_{\mu 0}\, q\,[\delta^3(\vec{x}-l\hat{e}) - \delta^3(\vec{x}+l\hat{e})] \quad . \tag{2}$$

$\hat{e}$ is a unit vector. With fixed q and l, a variation of the action with respect to $A_\mu(x)$ will determine the field $F_{\mu\nu}(x)$. Then the energy of the system, $E(q,l)$, can be computed. If we now scale q and l by

$$\begin{aligned} q &\to \lambda^2 q \;, \\ l &\to \lambda l \;, \end{aligned} \tag{3}$$

we obtain

$$E(q,l) = \frac{1}{\lambda^3}E(\lambda^2 q, \lambda l) \rightarrow E(q,l) = l^3\epsilon(q/l^2) \quad . \tag{4}$$

The linear confining and Coulomb cases, respectively, are given by

$$\begin{aligned} E &\sim ql \;, \\ E &\sim q^2/l \;. \end{aligned} \tag{5}$$

Equation (5) tells us that the string tension goes as the square root of the coefficient of the Coulomb potential. For most applications to non-Abelian statics this implies that the string tension is proportional to the square root of the Casimir operator in the representation of the external sources.

We turn now to the non-Abelian large-N case. The force between the external sources is extracted as follows[6]:

$$F_{[r]}(l) = -\frac{d}{dl}\frac{1}{T}\left[\lim_{T\to\infty} \ln\left(\frac{1}{d_{[r]}}\langle \chi_{[r]}(W_{T,l})\rangle\right)\right] \quad . \tag{6}$$

In (6) $[r]$ denotes the representation, $d_{[r]}$ is its dimension, $\chi_{[r]}$ its character, $W_{T,l}$ is the parallel transporter (untraced Wilson loop operator) around a planar rectangle of length T and width l, and $\langle\cdots\rangle$ denotes the expectation value. At infinite N, $SU(N)$ is undistinguishable from $U(N)$ and therefore

$$\frac{1}{d_{[\rm adj]}}\chi_{[\rm adj]}(W) = \left|\frac{1}{d_{[\rm fun]}}\chi_{[\rm fun]}(W)\right|^2 \quad . \tag{7}$$

Because of factorization at infinite N Eq. (7) holds also for the expectation values and, hence

$$F_{[\rm adj]}(l) = 2F_{[\rm fun]}(l) \quad . \tag{8}$$

Equation (8) is correct independently of l. Therefore, if we have confinement, the string tension scales precisely as the coefficient of the Coulomb force, that is, as the Casimir operator and not its square root.

As a check on our reasoning we consider the approximation in which the interaction between sources is assumed to be given solely by the two-point

$\langle A_\mu(x) A_\nu(y) \rangle$ Green's function. A partial resummation of Feynman diagrams has been claimed to give a $1/p^4$ singularity in the infrared, leading to a linear potential (albeit without flux confinement) at large distances.[7] Since the diagrams which were summed were all planar one should obtain a result consistent with the large-N argument. The use of only the two-point function allows Abelian modeling. Now, however, the requirement of locality cannot be met, and, in fact, a p^{-4} singularity in the propagator corresponds to an additional $\partial_\mu{}^2$ in the $F_{\mu\nu}{}^2$ term of (1). This precisely alters the scaling argument in the correct way.

To conclude, local nonlinear effective-action models for confinement (generalized flux-tube models) and the large-N limit predict different charge dependences for the linear confining potential. This result, together with the fact that the gauge group center is not felt at infinite N, suggests that different confinement mechanisms may be operative at infinite N and at finite N.[8] If this were the case, then no decisive lesson regarding $N=3$ confinement could be obtained from studies of the large-N limit.

This work was supported by the U.S. Department of Energy under Grant No. DE-AC02-76ER02220.

[1]H. Pagels and E. Tomboulis, Nucl. Phys. B143, 485 (1978); S. L. Adler, Phys. Rev. D 23, 2905 (1981); S. L. Adler and T. Piran, Phys. Lett. 113B, 405 (1982); 117B, 91 (1982); see also Y. Nambu, *ibid.* 102B, 149 (1981).
[2]P. Hasenfratz and J. Kuti, Phys. Rep. 40C, 75 (1978), and references therein.
[3]Yu. M. Makeenko and A. A. Migdal, Nucl. Phys. B188, 269 (1981).
[4]T. Banks and A. Casher, Nucl. Phys. B167, 215 (1980). See also J. Greensite and M. B. Halpern, Phys. Rev. D (to be published).
[5]E. Lieb, Rev. Mod. Phys. 53, 603 (1981); (unpublished).
[6]K. Wilson, Phys. Rev. D 10, 2445 (1974).
[7]M. Baker, J. S. Ball, and F. Zachariasen, Nucl. Phys. B186, 531 (1981); B186, 560 (1981); S. Mandelstam, Phys. Rev. D 20, 3223 (1979).
[8]G. 't Hooft, Nucl. Phys. B138, 1 (1978).

Nuclear Physics B244 (1984) 469–491

CHIRAL SYMMETRY BREAKING IN COULOMB GAUGE QCD*

S. L. ADLER

Institute for Advanced Study, Princeton, New Jersey 08540, USA

A.C. DAVIS[1]

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 9EW, UK

Received 3 April 1984

We analyze chiral symmetry breaking in QCD in Coulomb gauge. Using the Ward identities, we derive the renormalized gap equation from the renormalized Dyson equation for the vector and axial-vector vertices. We work within the ladder approximation, in which the Bethe–Salpeter kernel is a sum of longitudinal and transverse terms, depending only on momentum transfer. This relates the chiral symmetry breaking parameters to the static quark potential. When transverse gluon exchange is neglected, our gap equation agrees in the infrared with that obtained by Amer et al. from a non-normal-ordered Coulomb gauge hamiltonian, while disagreeing with the gap equation obtained by Finger and Mandula using a normal-ordering prescription. The corrected gap equation leads to infrared-finite formulas for the effective quark and pion parameters, in which integrals for physical quantities converge for an infrared-singular confining potential $V_c \propto q^{-4}$; we present the results of a numerical solution in this case.

1. Introduction

Quantum chromodynamics (QCD) is widely accepted as the candidate theory of strong interactions. Asymptotic freedom allows a perturbative treatment of QCD at high energies [1], permitting the theory to be tested in such processes as e^+e^- annihilation and electroproduction. Unfortunately, the bound state spectrum of QCD is still an unsolved problem. Progress in this direction has been made via Monte Carlo techniques in lattice gauge theories. However, the inclusion of fermions on the lattice is still problematic, particularly for massless fermions [2].

A pertinent question to ask of QCD with massless fermions is whether chiral symmetry is spontaneously broken, and by what mechanism. Formal arguments using the 't Hooft anomaly conditions [3] indicate that chiral symmetry in QCD must be broken in the Nambu–Goldstone mode. However, these arguments are essentially kinematical in nature, and the detailed dynamical mechanism of chiral symmetry breaking in QCD remains elusive.

* Supported by Department of Energy grant no. DE-AC02-76ER02220.
[1] SERC (UK) Advanced Fellow.

A popular approach to the dynamics of chiral symmetry breaking is to write down a gap equation for the generated mass Σ, which is then discussed analytically in the linearized approximation [4]. Usually this is done in the Landau gauge, where there is no wave function renormalization, yielding a gap equation which is finite without renormalization. However, in Landau gauge it is difficult [4] to implement renormalization group corrections to the gluon exchange potential. An alternative approach, pioneered by Finger and Mandula [5], is to construct the gap equation in Coulomb gauge. One advantage of using Coulomb gauge, which motivated the work of ref. [5], is that the Coulomb propagator corrections give the complete QCD β-function. This permits a simple implementation of renormalization-group-improved perturbation theory. A further advantage of Coulomb gauge, which we will exploit in sect. 3 below, is that the gap equation can be derived by making the same approximation to the quark-antiquark Bethe–Salpeter kernel as is used in the phenomenological analysis [6] of charmonium energy levels. This permits the use of phenomenological static potentials in the study of chiral symmetry breaking. However, a nontrivial problem with the Coulomb gauge is that, in this gauge, the wave function renormalization is infinite, and the gap equation requires renormalization.

At first sight it may seem rather strange to use the non-Lorentz-covariant Coulomb gauge, which singles out a preferred rest frame, to study a problem involving massless quarks. However, as seen below, as a result of dynamical chiral symmetry breaking, the effective quasi-particle excitations which arise from solving the gap equation have a nonzero mass. Thus, a preferred rest frame is defined: the frame in which the quasi-particles are at rest. The existence of this preferred frame gives an *a postiori* justification for the use of Coulomb gauge in setting up the gap equation.

In constructing the Coulomb gauge gap equation for a pairing-type model, Finger and Mandula [5] proceed from a normal-ordered Coulomb gauge hamiltonian. An analogous Coulomb-gauge gap equation, based on a non-normal-ordered hamiltonian, has been studied by Amer et al. [7]. These pairing models are reviewed in sect. 2, using the equivalent formalisms of the Bogoliubov–Valatin transformation and the Dyson equation to derive the gap equation. For a pure Coulomb potential, the models of refs. [5] and [7] can be shown to correspond to the use of a renormalized and of an unrenormalized gap equation, respectively. However, for a general phenomenological potential, the two models have qualitatively different infrared behavior. For example, the gap equation of Amer et al. exists for a confining potential, whereas that of Finger and Mandula does not. Now renormalizations to remove ultraviolet divergences should not change the infrared behavior of a theory. Thus for a general phenomenological potential, the models of ref. [5] and ref. [7] do not simply correspond to the renormalized and unrenormalized versions of the same pairing model. There is clearly a paradox here: either the equations of Amer et al. do not, in general, give the correct unrenormalized theory, or the equations of Finger and Mandula do not in general give the correct renormalized theory.

Our primary aim in this paper is to find the correct form of the renormalized gap equation in Coulomb gauge. This is done in sect. 3, starting from the renormalized Dyson equation for the vector and axial-vector vertex parts. We proceed by making the ladder approximation to the quark-antiquark Bethe–Salpeter kernel, which introduces phenomenological potentials. Application of the Ward identities unambiguously yields the correct form of the renormalized gap equation. From this analysis we reach the following conclusions:

(i) The equations of Amer et al. give in general the correct unrenormalized equations. This corresponds to the fact that, since the color matrices are traceless, the interaction term in QCD does not require normal ordering. For potentials which do not lead to ultraviolet divergences, such as a pure confining potential $V_c \propto q^{-4}$, $V_T = 0$, the gap equation of ref. [7] is the correct one as it stands.

(ii) The renormalized gap equation differs from the unrenormalized one by a polynomial counterterm, as expected from the standard BPHZ renormalization algorithm. The Finger–Mandula normal-ordering prescription does not correspond to a polynomial counterterm. Thus, it is incorrect, except for the case of a pure Coulomb potential.

A second principal objective of this paper is to study the infrared properties of the corrected gap equation. In sect. 4 we summarize an analysis of pion properties in the pairing model given recently by Govaerts, Mandula and Weyers [8]. We show that, when used in conjunction with the unrenormalized or correctly renormalized gap equation, all physical quantities are infrared-finite. That is, even for an infrared singular confining potential $V_c \propto q^{-4}$, all physical quantities are given by infrared-convergent integrals as a result of detailed cancellations between singular terms in the integrands. Furthermore, the infrared cancellations are shown to arise in a general way from the operator structure of the infrared singularity in the interaction hamiltonian. This feature is illustrated by a numerical example in sect. 5, where the gap equation and the pion vertex equation are solved numerically for the case of a pure q^{-4} potential.

To recapitulate, the plan of this paper is as follows. In sect. 2 we review the pairing models of refs. [5] and [7] and examine alternative renormalization prescriptions. In sect. 3 we determine the correct form of the renormalized gap equation by studying the Dyson equation for the renormalized vertex part. We demonstrate in sect. 4 the infrared finiteness of physical quantities when the correct gap equation is used. In sect. 5 we numerically solve the gap equation for the case of a confining potential. Finally, we summarize our findings in sect. 6.

2. Coulomb gauge pairing model

There are two equivalent approaches to setting up the Coulomb gauge pairing model, both of which are used in the paper of Finger and Mandula. The first is to make a Bogoliubov–Valatin transformation to a vacuum containing a $q\bar{q}$ condensate.

This is then optimized variationally, to give an equation (the gap equation) for the condensate wave function $\Psi(p)$. The second is to use the Dyson equation for the quark propagator, in the Hartree approximation, to set up an equation for the quark proper self-energy $\Sigma(p)$. This is the gap equation rewritten in a different notation. In this section we outline both methods, following the notation of Finger and Mandula, but without normal-ordering the interaction term in the hamiltonian.

The Coulomb gauge effective hamiltonian for a single* quark flavor q is

$$H_{\text{eff}} = \bar{q}\gamma\cdot(-i\nabla)q - 2\pi\sum_a \bar{q}\gamma_0\tfrac{1}{2}\lambda^a q\frac{\alpha}{\nabla^2}\bar{q}\gamma_0\tfrac{1}{2}\lambda^a q\,. \tag{2.1}$$

The summation is over the color index a, and $\alpha = g^2/4\pi$ can be taken to be either a constant or (as we shall do below) a running coupling $\alpha = \alpha(-\nabla^2)$. We wish to minimize H_{eff} over trial states containing a coherent superposition of $q\bar{q}$ pairs. The trial wave function is taken to be

$$|\Psi\rangle = \frac{1}{N(\Psi)}\prod_{p,s,a}[1 - s\Psi(p)\tau b^{(a)\dagger}(\boldsymbol{p},s)d^{(a)\dagger}(-\boldsymbol{p},s)]|0\rangle\,, \qquad p = |\boldsymbol{p}|\,, \tag{2.2}$$

with $b^{(a)\dagger}(\boldsymbol{p},s)$ and $d^{(a)\dagger}(\boldsymbol{p},s)$ the creation operators for a quark and antiquark with momentum $\boldsymbol{p}$, color index a and helicity s. In eq. (2.2) τ is the volume of an elementary cell in momentum space, and $\Psi(p) = \Psi^*(p)$ is the momentum-space pair wave function. The normalization factor $N(\Psi)$ is determined such that

$$(\Psi|\Psi\rangle = 1\,, \tag{2.3a}$$

to give

$$N(\Psi) = \prod_{p,s,a}\sqrt{1+\Psi(p)^2}\,. \tag{2.3b}$$

The calculation of matrix elements in the vacuum state $|\Psi\rangle$ is facilitated by making a Bogoliubov–Valatin transformation to an operator basis which annihilates $|\Psi\rangle$, as described in appendix A. For the equal time quark Feynman propagator we obtain

$$\begin{aligned}(\Psi|\tfrac{1}{2}[q_\alpha(\boldsymbol{x},0),\bar{q}_\beta(\boldsymbol{y},0)]|\Psi\rangle &= \int\frac{d^3p}{(2\pi)^3}e^{i\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{y})}\int\frac{dp_0}{2\pi}S^{(4)}(\boldsymbol{p},p_0)\\ &= \int\frac{d^3p}{(2\pi)^3}e^{i\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{y})}\left[\frac{\Psi(p)}{1+\Psi^2(p)} - \frac{1}{2}\frac{1-\Psi^2(p)}{1+\Psi^2(p)}\gamma\cdot\hat{p}\right]_{\alpha\beta},\end{aligned} \tag{2.4}$$

where $\hat{p}$ is the unit vector $\boldsymbol{p}/|\boldsymbol{p}|$. This enables us to obtain the following expression

* The generalization of the analysis to more than one quark flavor is easy.

for the expectation of H_{eff} in the state $|\Psi\rangle$,

$$\frac{\langle\Psi|H_{\text{eff}}|\Psi\rangle}{\delta^3(0)}=\int \mathrm{d}^3p\,\frac{12p\Psi^2(p)}{1+\Psi^2(p)}$$

$$-\frac{4}{\pi^2}\int \mathrm{d}^3p\,\mathrm{d}^3q\,\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\left[\frac{\Psi(p)\Psi(q)}{[1+\Psi^2(p)][1+\Psi^2(q)]}-\frac{1}{4}\right.$$

$$\left.+\left(\frac{\Psi^2(p)}{1+\Psi^2(p)}-\frac{1}{2}\right)\left(\frac{\Psi^2(q)}{1+\Psi^2(q)}-\frac{1}{2}\right)\hat{p}\cdot\hat{q}\right]. \tag{2.5}$$

The optimal condensate vacuum is obtained by minimizing $\langle\Psi|H|\Psi\rangle$ with respect to $\Psi(p)$,

$$\frac{\delta}{\delta\Psi(p)}\langle\Psi|H_{\text{eff}}|\Psi\rangle=0\,, \tag{2.6a}$$

which gives the "gap equation"

$$p\Psi(p)=\frac{1}{3\pi^2}\int \mathrm{d}^3q\,\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\,\frac{\Psi(q)[1-\Psi^2(p)]-\Psi(p)[1-\Psi^2(q)]\hat{p}\cdot\hat{q}}{1+\Psi^2(q)}. \tag{2.6b}$$

An alternative derivation of the gap equation is obtained from the Dyson equation for the quark propagator. Making a non-relativistic ansatz for the proper self-energy part Σ,

$$\Sigma=\Sigma(\boldsymbol{p})=pA(p)+\boldsymbol{\gamma}\cdot\boldsymbol{p}B(p)\,, \tag{2.7}$$

the propagator can be written as

$$S^{(4)}(\boldsymbol{p},p_0)=\frac{1}{\gamma_0p_0-\boldsymbol{\gamma}\cdot\boldsymbol{p}-\Sigma}$$

$$=\frac{\gamma_0p_0-\boldsymbol{\gamma}\cdot\boldsymbol{p}[1+B(p)]+pA(p)}{p_0^2-\omega(p)^2}\,,$$

$$\omega(p)=p\sqrt{A^2(p)+[1+B(p)]^2}\,. \tag{2.8}$$

After integration over p_0 this gives

$$S^{(3)}(\boldsymbol{p})=\int\frac{\mathrm{d}p_0}{2\pi}\,iS^{(4)}(\boldsymbol{p},p_0)=\frac{pA(p)-p[1+B(p)]\gamma\cdot\hat{p}}{2\omega(p)}\,, \tag{2.9}$$

and comparing eqs. (2.8) and (2.9) with eq. (2.4) we see that

$$\frac{A(p)}{\sqrt{A^2(p)+[1+B(p)]^2}}=\frac{2\Psi(p)}{1+\Psi^2(p)}\equiv\sin 2\vartheta(p)\,,$$

$$\frac{1+B(p)}{\sqrt{A^2(p)+[1+B(p)]^2}}=\frac{1-\Psi^2(p)}{1+\Psi^2(p)}\equiv\cos 2\vartheta(p)\,, \tag{2.10}$$

with $\vartheta(p)$ the rotation angle of the Bogoliubov–Valatin transformation defined in eq. (A.4). In the Hartree approximation the Dyson equation for Σ is

$$\Sigma(p)=\frac{2}{3\pi^2}\int \mathrm{d}^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\gamma_0 S^{(3)}(\boldsymbol{q})\gamma_0\,, \tag{2.11}$$

which, on substitution of eq. (2.9), yields separate equations for $A(p)$ and $B(p)$,

$$pA(p)=\frac{2}{3\pi^2}\int \mathrm{d}^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\frac{\Psi(q)}{1+\Psi^2(q)}\,, \tag{2.12a}$$

$$pB(p)=\frac{2}{3\pi^2}\int \mathrm{d}^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\frac{1}{2}\frac{1-\Psi^2(q)}{1+\Psi^2(q)}\hat{p}\cdot\hat{q}\,. \tag{2.12b}$$

When eqs. (2.12a) and (2.12b) are substituted into the identity

$$p\Psi(p)=\tfrac{1}{2}[1-\Psi^2(p)]pA(p)-\Psi(p)pB(p)\,, \tag{2.13}$$

we obtain again the gap equation of eq. (2.6b).

Eqs. (2.1)–(2.13) are equivalent to the analysis of Amer et al. [7] who, as discussed in sect. 1, proceed from the non-normal-ordered Coulomb gauge hamiltonian. The calculation of Finger and Mandula [5] instead starts from a hamiltonian analogous to that of eq. (2.1), but normal-ordered with respect to the perturbative vacuum $|0\rangle$, which yields the following results. For the gap equation, Finger and Mandula obtain

$$p\Psi(p)=\frac{1}{3\pi^2}\int \mathrm{d}^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\frac{\Psi(q)[1-\Psi^2(p)]+2\hat{p}\cdot\hat{q}\Psi(p)\Psi^2(q)}{1+\Psi^2(q)}\,, \tag{2.6b$'$}$$

which, in the Green function approach, corresponds to the "Dyson-like" equation

$$\Sigma(p)=\frac{2}{3\pi^2}\int \mathrm{d}^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\gamma_0[S^{(3)}(\boldsymbol{q})-S_0^{(3)}(\boldsymbol{q})]\gamma_0\,, \tag{2.11$'$}$$

with S_0^3 the free zero-mass propagator

$$S_0^{(3)}(\boldsymbol{p})^{-1}=-\tfrac{1}{2}\boldsymbol{\gamma}\cdot\hat{\boldsymbol{p}}\,. \tag{2.14}$$

Since pairing is a low-momentum effect, the condensate wave function Ψ is expected to vanish rapidly at high momenta,

$$\Psi(p)\xrightarrow[p\to\infty]{}0\,. \tag{2.15}$$

Consequently eq. (2.11) and eq. (2.11$'$) exhibit very different high-momentum behavior. For a Coulomb potential

$$\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)=\text{const}\,, \tag{2.16a}$$

or for an asymptotically free Coulomb-like potential

$$\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)=\text{const}/\log[|\boldsymbol{p}-\boldsymbol{q}|^2/\Lambda^2]\,, \tag{2.16b}$$

the integral in eq. (2.6b′) converges at high momenta as a result of the overall factor $\Psi(q)$ in the integrand, while the integral in eq. (2.6b) has a high-momentum divergence given by

$$\text{divergent part of eq. (2.6b)} = (Z-1)p\Psi(p)\,,$$

where

$$Z-1=\frac{1}{3\pi^2}\int d^3q\frac{d}{dq^2}\left[\frac{\alpha(q^2)}{q^2}\right]\tfrac{2}{3}q\,. \tag{2.17}$$

Hence, the Finger–Mandula gap equation is ultraviolet-finite, while that of Amer et al. has an ultraviolet divergence given by eq. (2.17). However, since the divergence in the Amer et al. equation is proportional to the kinetic term on the left-hand side of the gap equation, it can be eliminated by the standard renormalization procedure of adding a counterterm

$$\Delta H=(Z-1)\bar{q}\gamma\cdot(-i\nabla)q \tag{2.18}$$

to the hamiltonian of eq. (2.1). This gives an alternative subtraction scheme to the one used by Finger and Mandula. Taking the difference between eq. (2.6b′) and eq. (2.6b) gives

$$\text{eq. (2.6b′)–eq. (2.6b)} = [Z(p)-1]p\Psi(p)\,,$$

$$Z(p)-1\equiv\frac{1}{p}\frac{1}{3\pi^2}\int d^3q\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2}\hat{p}\cdot\hat{q}\,. \tag{2.19}$$

Thus, the two alternative schemes are equivalent only in the case of a pure Coulomb potential (eq. (2.16a)), for which $Z(p)$ in eq. (2.19) reduces to a constant. That the Finger–Mandula subtraction scheme corresponds to use of a momentum-dependent renormalization constant leads one to suspect that it is incorrect. This conclusion is confirmed in the next section, where eqs. (2.1)–(2.13) and (2.18) are derived, via the Ward identities, from the Dyson equations for the renormalized vertex parts.

3. The renormalized Coulomb gauge gap equation*

In the theory of superconductivity, the gap equation is part of a more general system of equations for the electron propagator and the electron–photon vertex part [9], in which the Ward identity is satisfied. Hence, in choosing between alternative subtraction schemes for the pairing model of sect. 2, it is natural to proceed in an analogous fashion, using the Ward identities to derive the renormalized gap equation from suitable approximations to the vertex parts. Specifically, we start from the renormalized Dyson equation for the vector and axial-vector vertices, making the

* In this section, we use q and p to denote four-vectors, whereas in all other sections of this paper, the momenta q and p denote the three-vector magnitudes $q=|\boldsymbol{q}|$, $p=|\boldsymbol{p}|$.

ladder approximation that the Bethe–Salpeter kernel depends only on the momentum transfer. Consistent with this approximation, we exclude quark annihilation graphs, so that there are no anomalies, and neglect terms arising from the non-commutativity of color matrices on the quark lines, so that we can use the Ward identities of QED (rather than the more complicated Slavnov–Taylor identities of QCD) and can omit all color indices.

Our analysis therefore proceeds from the following equations for the renormalized vector and axial-vector vertex parts [10],

$$\tilde{\Gamma}_\mu(p',p)_{\delta\gamma} = (Z\gamma_\mu)_{\delta\gamma} + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}'_F(p'+q)\tilde{\Gamma}_\mu(p'+q,p+q)i\tilde{S}'_F(p+q)]_{\beta\alpha}$$
$$\times \tilde{K}_{\alpha\beta,\gamma\delta}(p+q,p'+q,q)\,,$$
$$\tilde{\Gamma}_{\mu5}(p',p)_{\delta\gamma} = [(Z\gamma_\mu)\gamma_5]_{\delta\gamma} + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}'_F(p'+q)\tilde{\Gamma}_{\mu5}(p'+q,p+q)i\tilde{S}'_F(p+q)]_{\beta\alpha}$$
$$\times \tilde{K}_{\alpha\beta,\gamma\delta}(p+q,p'+q,q)\,. \tag{3.1}$$

In eq. (3.1) $\tilde{S}'_F$ is the full renormalized quark propagator, $\tilde{K}_{\alpha\beta,\gamma\delta}$ is the renormalized quark-antiquark Bethe–Salpeter kernel, and $(Z\gamma_\mu)$ is a shorthand for

$$(Z\gamma_\mu) = \begin{cases} Z_0\gamma_0, & \mu=0 \\ Z\gamma_j, & \mu=j=1,2,3. \end{cases} \tag{3.2}$$

We are thus anticipating the fact that, if we make a non-covariant approximation to the ultraviolet tail of $\tilde{K}$, the $\mu=0$ and $\mu=1,2,3$ components of the vertex can have different renormalizations. Since anomalies have been excluded, the vector and axial-vector vertex parts have the same renormalization constants [11], and satisfy the Ward identities

$$(p'-p)^\mu\tilde{\Gamma}_\mu(p',p) = \tilde{S}'^{-1}_F(p') - \tilde{S}'^{-1}_F(p)\,,$$
$$(p'-p)^\mu\tilde{\Gamma}_{\mu5}(p',p) = \gamma_5\tilde{S}'^{-1}_F(p) + \tilde{S}'^{-1}_F(p')\gamma_5\,. \tag{3.3}$$

Let us now make our fundamental approximation. This is to assume that the Bethe–Salpeter kernel $\tilde{K}_{\alpha\beta,\gamma\delta}(p+q,p'+q,q)$ is a function only of the momentum transfer q, with Lorentz vector couplings* on the quark and antiquark lines,

$$\tilde{K}_{\alpha\beta,\gamma\delta}(p+q,p'+q,q) \approx \tilde{k}_{\alpha\beta,\gamma\delta}(q)$$
$$= -4\pi i(\gamma_0)_{\delta\beta}(\gamma_0)_{\alpha\gamma}\tfrac{4}{3}V_C(|\boldsymbol{q}|)$$
$$-4\pi i(\gamma_j)_{\delta\beta}(\gamma_k)_{\alpha\gamma}\left(\delta_{jk}-\frac{q_jq_k}{\boldsymbol{q}^2}\right)\tfrac{4}{3}V_T(q)\,. \tag{3.4}$$

* The introduction of scalar couplings would lead to explicit (as opposed to spontaneous dynamical) breakdown of chiral symmetry.

This approximation is the usual one made in potential theory treatments of heavy quark bound states. In lowest-order perturbation theory, V_C and V_T are given respectively by single Coulomb and single transverse gluon exchange,

$$V_C = \frac{\alpha}{\boldsymbol{q}^2}, \qquad V_T = \frac{\alpha}{q_0^2 - \boldsymbol{q}^2}. \tag{3.5}$$

In higher loop order, renormalization group improvements make α in V_C a running function of $\boldsymbol{q}$ at high momentum, while multiple Coulomb gluon exchange [12] produces an infrared-singular confining contribution to V_C at low momentum. For example, corresponding to the frequently used [6] coordinate space potential

$$V(r) = \sigma r - \frac{4}{3}\frac{\alpha}{r}, \tag{3.6}$$

the momentum-space potential (defined as the Fourier transform of $V(r)$ with a factor $-\frac{16}{3}\pi$ divided out) would be

$$V_C(|\boldsymbol{q}|) = \frac{\alpha}{\boldsymbol{q}^2} + \frac{2\kappa}{(\boldsymbol{q}^2)^2}, \qquad \kappa \equiv \tfrac{3}{4}\sigma, \tag{3.7}$$

with α in eqs. (3.6)–(3.7) either a constant or a running logarithmic function of the coordinate or momentum, and with the term proportional to κ the confining potential.

Before using specific assumptions about the form of V_C and V_T, we first carry the analysis as far as possible using only the approximation to the form of $\tilde{K}$ given in eq. (3.4)*. Substituting eq. (3.4) into eq. (3.1), eq. (3.1) into the Ward identities of eq. (3.3), and using the Ward identities a second time to rearrange the integrands, we obtain from the vector vertex equation

$$\begin{aligned}
(p'-p)^\mu \tilde{\Gamma}_\mu(p',p) &= \tilde{S}_F'^{-1}(p') - \tilde{S}_F'^{-1}(p) \\
&= (Z\gamma_\mu)(p'-p)^\mu + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}_F'(p'+q) \\
&\quad \times (\tilde{S}_F'^{-1}(p'+q) - \tilde{S}_F'^{-1}(p+q))i\tilde{S}_F'(p+q)]_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q) \\
&= (Z\gamma_\mu)p'^\mu + \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p'+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q) \\
&\quad - (Z\gamma_\mu)p^\mu - \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q).
\end{aligned} \tag{3.8}$$

* The derivation of eqs. (3.8)–(3.12) is in fact independent of the choice of gauge, since it uses only the assumptions that $\tilde{k}$ depends solely on the momentum transfer q and has vector couplings on the fermion lines.

Similarly, from the axial-vector vertex equation we get

$$\begin{aligned}
(p'-p)^\mu \tilde{\Gamma}_{\mu 5}(p',p) &= \gamma_5 \tilde{S}_F'^{-1}(p) + \tilde{S}_F'^{-1}(p')\gamma_5 \\
&= (Z\gamma_\mu\gamma_5)(p'-p)^\mu + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}_F'(p'+q) \\
&\quad \times (\gamma_5 \tilde{S}_F'^{-1}(p+q) + \tilde{S}_F'^{-1}(p'+q)\gamma_5) i\tilde{S}_F'(p+q)]_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q) \\
&= \gamma_5\left[(Z\gamma_\mu)p^\mu + \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q)\right] \\
&\quad + \left[(Z\gamma_\mu)p'^\mu + \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p'+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q)\right]\gamma_5\,,
\end{aligned} \tag{3.9}$$

where the Lorentz vector structure of $\tilde{k}$ has been used in anticommuting the γ_5's to the outside on the right-hand side. Because the p and p' dependence in eqs. (3.8) and (3.9) has separated, we deduce that the renormalized propagator must satisfy

$$\tilde{S}_F'^{-1}(p) = (Z\gamma_\mu)p^\mu + \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\dots}(q)\,. \tag{3.10}$$

Note that the presence of an additive constant (a kinematical mass term) in eq. (3.10), which would be allowed by the vector Ward identity of eq. (3.8), is excluded by the axial-vector Ward identity of eq. (3.9)! Introducing the self-energy Σ by writing

$$\tilde{S}_F'^{-1}(p) = \gamma_\mu p^\mu - \Sigma(p)\,, \tag{3.11}$$

we see that Σ must satisfy the integral equation

$$\Sigma(p)_{\delta\gamma} = [\gamma_\mu - (Z\gamma_\mu)]_{\delta\gamma}p^\mu - \int \frac{d^4q}{(2\pi)^4}\tilde{S}_F'(p+q)_{\beta\alpha}\tilde{k}_{\alpha\beta,\gamma\delta}(q)\,. \tag{3.12}$$

This is the renormalized gap equation corresponding to the ladder approximation to the Bethe–Salpeter kernel.

To make contact with the non-relativistic pairing model of sect. 2, let us now make the further approximation of neglecting the transverse gluon exchange term in $\tilde{k}$. Thus, $\tilde{k}$ becomes

$$\tilde{k}_{\alpha\beta,\gamma\delta}(q) = -4\pi i(\gamma_0)_{\delta\beta}(\gamma_0)_{\alpha\gamma}\tfrac{4}{3}V_C(|\boldsymbol{q}|)\,. \tag{3.13a}$$

Further, let us assume that $\Sigma(p)$ is given by the non-relativistic ansatz

$$\Sigma = \Sigma(\boldsymbol{p}) = |\boldsymbol{p}|A(|\boldsymbol{p}|) + \boldsymbol{\gamma}\cdot\boldsymbol{p}B(|\boldsymbol{p}|)\,. \tag{3.13b}$$

When eq. (3.13) is substituted into the integral equation for the $\mu=0$ component of the vector vertex at zero momentum transfer, this equation simplifies dramatically. It is solved by

$$\tilde{\Gamma}_0(p,p) = Z_0\gamma_0\,. \tag{3.14}$$

To verify this, we note that when both eqs. (3.13) and (3.14) are substituted into eq. (3.1), the q-integral

$$\int \frac{d^4q}{(2\pi)^4}[i\tilde{S}'_F(p+q)\tilde{\Gamma}_0(p+q,p+q)i\tilde{S}'_F(p+q)]_{\beta\alpha}\tilde{K}_{\alpha\beta,\gamma\delta}(p+q,p+q,q) \tag{3.15}$$

reduces to

$$\frac{i}{2\pi^2}\int d^3q\, V_C(|\boldsymbol{q}|)I\,,$$

$$I=\int_{-\infty}^{\infty}\frac{dq_0}{2\pi}\,\gamma_0\tilde{S}'_F\gamma_0\tilde{S}'_F\gamma_0\,,$$

$$\tilde{S}'_F=[\gamma_0(q_0+p_0)-\boldsymbol{\gamma}\cdot(\boldsymbol{q}+\boldsymbol{p})(1+B)-|\boldsymbol{q}+\boldsymbol{p}|A]^{-1}\,. \tag{3.16}$$

Evaluation of the q_0-integral shows that $I\equiv 0$. Since asymptotic freedom requires $\tilde{\Gamma}_0(p,p)$ to approach γ_0 at large momentum, we conclude that, in the Coulomb gluon exchange model,

$$Z_0=1\,. \tag{3.17}$$

Substituting eqs. (3.13) and (3.17) into eq. (3.12), the gap equation then simplifies to

$$\Sigma(\boldsymbol{p})=(Z-1)\boldsymbol{\gamma}\cdot\boldsymbol{p}+\frac{2}{3\pi^2}\int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|)\gamma_0 S^{(3)}(\boldsymbol{q})\gamma_0\,, \tag{3.18}$$

which, when reexpressed in terms of Ψ by following the analysis of eqs. (2.10)–(2.13), takes the form

$$|\boldsymbol{p}|\Psi(|\boldsymbol{p}|)=(1-Z)|\boldsymbol{p}|\Psi(|\boldsymbol{p}|)$$
$$+\frac{1}{3\pi^2}\int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|)\frac{\Psi(|\boldsymbol{q}|)[1-\Psi^2(|\boldsymbol{p}|)]-\Psi(|\boldsymbol{p}|)[1-\Psi^2(|\boldsymbol{q}|)]\hat{p}\cdot\hat{q}}{1+\Psi^2(|\boldsymbol{q}|)}\,. \tag{3.19}$$

When the obvious identification

$$V_C(|\boldsymbol{p}-\boldsymbol{q}|)=\frac{\alpha(|\boldsymbol{p}-\boldsymbol{q}|^2)}{|\boldsymbol{p}-\boldsymbol{q}|^2} \tag{3.20}$$

is made, eq. (3.18) (eq. (3.19)) is identical in structure to the gap equation of eq. (2.11) (eq. (2.6b)), apart from the addition of a renormalization counterterm with precisely the form derived from ΔH of eq. (2.18). Thus, our rederivation of the gap equation from an analysis of the vertex parts shows that the standard renormalization algorithm, rather than the Finger–Mandula normal-ordering prescription, gives the correct method for eliminating ultraviolet divergences.

4. Infrared finiteness of physical quantities

Having established the correct renormalization prescription, let us now consider the infrared behavior of the gap equation. As mentioned briefly in sect. 1, the Finger–Mandula gap equation (eq. (2.6b′)) has very different infrared behavior from that of the Amer et al. unrenormalized gap equation* (eq. (2.6b)) and of the correctly renormalized gap equation (eq. (3.19)). Specifically, as $\boldsymbol{p} \to \boldsymbol{q}$ the integrand of eq. (2.6b′) simplifies to

$$V_C(|\boldsymbol{p}-\boldsymbol{q}|)\,\Psi(p)\,, \tag{4.1a}$$

while the integrand of eqs. (2.6b) and (3.19) reduces to

$$V_C(|\boldsymbol{p}-\boldsymbol{q}|)[\hat{\boldsymbol{p}}\cdot(\boldsymbol{q}-\boldsymbol{p})\,\Psi'(p)+\mathrm{O}((\boldsymbol{q}-\boldsymbol{p})^2)]\,. \tag{4.2a}$$

Hence, when V_C is an infrared-singular confining potential diverging as $|\boldsymbol{p}-\boldsymbol{q}|^{-4}$ as $\boldsymbol{p} \to \boldsymbol{q}$, the q-integral in eq. (2.6b′) behaves as

$$\int \mathrm{d}^3q|\boldsymbol{p}-\boldsymbol{q}|^{-4}\,\Psi(p) \tag{4.1b}$$

and diverges, while the corresponding integral in eqs. (2.6b) and (3.19) behaves, after angular averaging, as

$$\int \mathrm{d}^3q|\boldsymbol{p}-\boldsymbol{q}|^{-4}\mathrm{O}((\boldsymbol{q}-\boldsymbol{p})^2) \tag{4.2b}$$

and converges. *Thus, the corrected gap equation and the pair wave function Ψ exist for a confining potential.* We will refer to Ψ, and to any other quantity in the pairing model which exists for a confining potential, as being "infrared-finite."

A second feature of the correctly renormalized gap equation, related to its infrared-finiteness, is that it is invariant in form when the coordinate-space potential is shifted by a uniform constant,

$$V(r) \to V(r) - \tfrac{16}{3}\pi C\,. \tag{4.3a}$$

In terms of the momentum space potential of eq. (3.7), the transformation of eq. (4.3a) takes the form

$$V_C(|\boldsymbol{p}-\boldsymbol{q}|) \to V_C(|\boldsymbol{p}-\boldsymbol{q}|) + (2\pi)^3 C\delta^3(\boldsymbol{p}-\boldsymbol{q})\,. \tag{4.3b}$$

This is obviously an invariance of eqs. (2.6b), (3.19) and (4.2a), since the coefficient of $V_C(|\boldsymbol{p}-\boldsymbol{q}|)$ vanishes at $\boldsymbol{p}=\boldsymbol{q}$ in these equations. For reasons explained below, we expect physical observables very generally to be invariant under the transformation of eq. (4.3). This, in turn, means that in the formulas for all physical observables, V_C must appear multiplied by a coefficient which vanishes at $\boldsymbol{p}=\boldsymbol{q}$. So by the angular

* Strictly speaking, Amer et al. [7] give only the linearized form of eq. (2.6b). The full gap equation, and the observation that is infrared-finite, appear in a subsequent paper by Le Yaouanc et al. [13].

averaging argument used in eq. (4.2), we conclude that *all physical observables are infrared finite.* In the remainder of this section, we will illustrate this general statement by an explicit, case-by-case examination of various physical observables which can be formed in the pairing model.

Let us begin with the quasi-particle energy $\omega(p)$. According to eqs. (2.8), (2.10), (2.12a) and (3.20), this can be written as

$$\omega(p)=\frac{pA(p)}{\sin 2\vartheta(p)}=\frac{1}{3\pi^2}\int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|)\frac{\sin 2\vartheta(q)}{\sin 2\vartheta(p)}$$
$$=\frac{1}{3\pi^2}\int d^3q\, V_C(q)\frac{\sin 2\vartheta(|\boldsymbol{p}-\boldsymbol{q}|)}{\sin 2\vartheta(p)}. \tag{4.4}$$

Under the shift in potential of eq. (4.3b), $\omega(p)$ changes to

$$\omega(p)\to\omega(p)+\tfrac{8}{3}\pi C. \tag{4.5}$$

Hence the quasi-particle energy is not a physical observable and is clearly not infrared-finite*. However, since the shift in ω in eq. (4.5) is a constant, the *excitation energy* $\omega(p)-\omega(0)$ is a physical observable, and is given by either of the two equivalent infrared-finite formulas,

$$\omega(p)-\omega(0)=\frac{1}{3\pi^2}\int d^3q\, V_C(q)\left[\frac{\sin 2\vartheta(|\boldsymbol{p}-\boldsymbol{q}|)}{\sin 2\vartheta(p)}-\sin 2\vartheta(q)\right], \tag{4.6a}$$
$$\omega(p)-\omega(0)=\tilde{\omega}(p)-\tilde{\omega}(0),$$
$$\tilde{\omega}(p)=\frac{1}{3\pi^2}\int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|)\left[\frac{\sin 2\vartheta(q)}{\sin 2\vartheta(p)}-1\right]. \tag{4.6b}$$

In writing eq. (4.6a), we have used the fact that the gap equation of eq. (3.19) implies that

$$\Psi(0)=1\Rightarrow\sin 2\vartheta(0)=1\,. \tag{4.7}$$

The formula in eq. (4.6b) has advantages in numerical work, since $\sin 2\vartheta$ appears only in the radial integral over q, not in the angular integral over $\hat{q}$.

Next we consider the quark condensate $\langle\bar{u}u\rangle$ for a single quark flavor u. Following ref. [5], this can be evaluated in terms of the propagator $S^{(3)}(\boldsymbol{q})$,

$$\langle\bar{u}u\rangle=\langle\Psi|\bar{q}q|\Psi\rangle=3\delta_{\alpha\beta}\langle\Psi|\tfrac{1}{2}[\bar{q}_\beta(0),q_\alpha(0)]+\tfrac{1}{2}\{\bar{q}_\beta(0),q_\alpha(0)\}|\Psi\rangle$$
$$=-3\int\frac{d^3p}{(2\pi)^3}\mathrm{Tr}\,S^{(3)}(\boldsymbol{p})=-3\int\frac{d^3p}{(2\pi)^3}2\frac{pA(p)}{\omega(p)}, \tag{4.8}$$

where the color factor, 3, arises from the trace over the suppressed color index. Although $\omega(p)$ and $pA(p)$ are individually not physical observables and are not

* We thus again differ here with Finger and Mandula, who interpret $\omega(0)$ as the quasi-particle mass, and circumvent the infrared finiteness problem by using only potentials which are cut off in the infrared.

infrared-finite, eqs. (2.8) and (2.10) show that the ratio $pA(p)/\omega(p)$ can be expressed entirely in terms of the condensate wave function Ψ, and so is infrared-finite,

$$\frac{pA(p)}{\omega(p)}=\frac{2\Psi(p)}{1+\Psi^2(p)}=\sin 2\vartheta(p)\,,$$

$$\langle\bar{u}u\rangle=-\frac{3}{\pi^2}\int_0^\infty p^2\,\mathrm{d}p\,\sin 2\vartheta(p)\,. \tag{4.9}$$

Finally, we turn to pion properties in the pairing model. These have been investigated by Govaerts, Mandula and Weyers [8], who use the Bethe-Salpeter equation, with the approximated kernel of eq. (3.13a), to compute static pion properties in the chiral symmetry limit. Their algebraic calculation is unaffected by changing the subtraction scheme for the gap equation from eq. (2.6b′) to eq. (3.19), so we only quote their final results. These are summarized by the formulas

$$\frac{N\mu_\pi^2 f_\pi}{2m_q}=\langle\bar{u}u\rangle\,, \tag{4.10a}$$

$$-Nf_\pi=N^2=3\int\frac{\mathrm{d}^3p}{(2\pi)^3}\sin 2\vartheta(p)\frac{\bar{P}(p)}{\omega^2(p)}\,. \tag{4.10b}$$

In eq. (4.10) μ_π is the pion mass, m_q the constituent quark mass, f_π the pion decay constant, and N the normalization of the pion Bethe-Salpeter wave function. The pion vertex part form factor, $\bar{P}(p)$, satisfies the integral equation

$$\bar{P}(p)=pA(p)+\frac{1}{3\pi^2}\int\mathrm{d}^3q\frac{V_{\mathrm{C}}(|\boldsymbol{p}-\boldsymbol{q}|)}{\omega^3(q)}[pA(p)qA(q)+\boldsymbol{p}\cdot\boldsymbol{q}C(p)C(q)]\bar{P}(q)\,,$$

$$C(p)=1+B(p)\,. \tag{4.11}$$

Introducing the abbreviated notation

$$g(p)=\frac{\bar{P}(p)}{\omega^2(p)}\,, \tag{4.12}$$

Eq. (4.10b) is equivalent to

$$-N=f_\pi=\left[\frac{3}{2\pi^2}\int_0^\infty p^2\,\mathrm{d}p\,\sin 2\vartheta(p)g(p)\right]^{1/2}\,, \tag{4.13}$$

and will be infrared-finite if $g(p)$ is infrared-finite. Substituting eq. (4.12) into eq. (4.11), dividing by $\omega(p)$ and using eq. (2.10), the equation determining $g(p)$ can be written as

$$g(p)\omega(p)=\sin 2\vartheta(p)+\frac{1}{3\pi^2}\int\mathrm{d}^3q\,V_{\mathrm{C}}(|\boldsymbol{p}-\boldsymbol{q}|)$$

$$\times[\sin 2\vartheta(p)\sin 2\vartheta(q)+\hat{p}\cdot\hat{q}\cos 2\vartheta(p)\cos 2\vartheta(q)]g(q)\,. \tag{4.14}$$

To show that eq. (4.14) is infrared-finite, we express the unrenormalized* gap equation of eq. (2.6b) in terms of $\vartheta(p)$,

$$p \sin 2\vartheta(p) = \frac{1}{3\pi^2} \int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|) \times[\sin 2\vartheta(q) \cos 2\vartheta(p) - \hat{p}\cdot\hat{q} \cos 2\vartheta(q) \sin 2\vartheta(p)], \quad (4.15)$$

and use this to rewrite eq. (4.4) as follows,

$$\begin{aligned}\omega(p) &= \frac{1}{3\pi^2} \int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|) \sin 2\vartheta(q) \left(\sin 2\vartheta(p) + \frac{\cos^2 2\vartheta(p)}{\sin 2\vartheta(p)}\right) \\ &\quad + \frac{\cos 2\vartheta(p)}{\sin 2\vartheta(p)} \Big\{ p \sin 2\vartheta(p) - \frac{1}{3\pi^2} \int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|) \\ &\quad \times[\sin 2\vartheta(q) \cos 2\vartheta(p) - \hat{p}\cdot\hat{q} \cos 2\vartheta(q) \sin 2\vartheta(p)]\Big\} \\ &= p \cos 2\vartheta(p) + \frac{1}{3\pi^2} \int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|) \\ &\quad \times[\sin 2\vartheta(p) \sin 2\vartheta(q) + \hat{p}\cdot\hat{q} \cos 2\vartheta(p) \cos 2\vartheta(q)]. \end{aligned} \quad (4.16)$$

Substituting eq. (4.16) into eq. (4.14), the equation determining $g(p)$ finally becomes

$$\begin{aligned} g(p)p \cos 2\vartheta(p) &= \sin 2\vartheta(p) + \frac{1}{3\pi^2} \int d^3q\, V_C(|\boldsymbol{p}-\boldsymbol{q}|) \\ &\quad \times[\sin 2\vartheta(p) \sin 2\vartheta(q) + \hat{p}\cdot\hat{q} \cos 2\vartheta(p) \cos 2\vartheta(q)] \\ &\quad \times[g(q) - g(p)], \end{aligned} \quad (4.17)$$

which is manifestly infrared-finite.

A useful insight into the infrared-divergence structure of the pairing model is obtained by writing the propagator of eq. (2.8) in the form

$$S^{(4)}(\boldsymbol{p}, p_0) = \frac{R_+(\boldsymbol{p})}{p_0 - \omega(0) - [\omega(p) - \omega(0)]} + \frac{R_-(\boldsymbol{p})}{p_0 + \omega(0) + [\omega(p) - \omega(0)]}, \quad (4.18a)$$

with the residues $R_\pm(\boldsymbol{p})$ given by the infrared-finite expressions

$$\begin{aligned} R_\pm(\boldsymbol{p}) &= \frac{1}{2}\left[\gamma_0 \pm \frac{pA(p)}{\omega(p)}\right] \mp \tfrac{1}{2}\boldsymbol{\gamma}\cdot\boldsymbol{p}\frac{1+B(p)}{\omega(p)} \\ &= \frac{1}{2}\left[\gamma_0 \pm \frac{2\Psi(p)}{1+\Psi^2(p)}\right] \mp \tfrac{1}{2}\boldsymbol{\gamma}\cdot\boldsymbol{p}\frac{1}{p}\frac{1-\Psi^2(p)}{1+\Psi^2(p)} \\ &= \tfrac{1}{2}[\gamma_0 \pm \sin 2\vartheta(p)] \mp \tfrac{1}{2}\boldsymbol{\gamma}\cdot\boldsymbol{p}\frac{\cos 2\vartheta(p)}{p}. \end{aligned} \quad (4.18b)$$

* Use of the renormalized gap equation of eq. (3.19) in the following analysis would add a polynomial counterterm to eqs. (4.16) and (4.17), but would not change their infrared structure.

The only infrared-divergent quantity in eq. (4.18) is the term $\omega(0)$ in the denominators. This structure is a reflection of the fact that as a result of confinement, an infinite amount of energy is required to create a single quasiparticle state from the vacuum.

The fact that the infrared divergences take the form of eq. (4.18) has a simple operator interpretation. To see this, we consider the extension of the hamiltonian of eq. (2.1) to a general phenomenological potential $V(r)$,

$$H=\int d^3x\,\bar{q}\boldsymbol{\gamma}\cdot(-i\boldsymbol{\nabla})q-\tfrac{1}{2}\int d^3x\,d^3y\,q^\dagger(x)\tfrac{1}{2}\lambda^a q(x)\tfrac{3}{4}V(|x-y|)q^\dagger(y)\tfrac{1}{2}\lambda^a q(y)\,. \tag{4.19a}$$

When the interaction term in eq. (4.19a) is rewritten in terms of the momentum space potential V_C of eq. (3.7), we get

$$H=\int d^3x\,\bar{q}\boldsymbol{\gamma}\cdot(-i\boldsymbol{\nabla})q+\frac{1}{(2\pi)^2}\int d^3q\,\rho^a(\boldsymbol{q})\rho^a(-\boldsymbol{q})V_C(|\boldsymbol{q}|)\,, \tag{4.19b}$$

with

$$\rho^a(\boldsymbol{q})=\int d^3x\,e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}q^\dagger(x)\tfrac{1}{2}\lambda^a q(x)\,. \tag{4.20}$$

Hence the infrared-divergent part of the interaction term has the operator structure

$$\frac{1}{(2\pi)^2}\int d^3q\,V_C(|\boldsymbol{q}|)\rho^a(0)\rho^a(0)\,, \tag{4.21}$$

and so is proportional to the color-squared operator $F^2\equiv\rho^a(0)\rho^a(0)$. Similarly, under the shift in potential of eq. (4.3), the change in the hamiltonian is

$$H\to H+2\pi CF^2\,, \tag{4.22}$$

and is again proportional to the color-squared operator. Because color-squared is governed by a superselection rule, any eigenstate ψ of H with energy E and color-squared $C_2(\psi)$ is also an eigenstate of the shifted H of eq. (4.22), with the energy shifted to $E+2\pi CC_2(\psi)$. By the same reasoning, the linear infrared divergence of eq. (4.21) appears solely as an energy level shift

$$\frac{1}{(2\pi)^2}\int d^3q\,V_C(|\boldsymbol{q}|)C_2(\psi)\,. \tag{4.23}$$

The contribution of this term to the energy difference $\omega(0)$ between the single quasi-particle state ($C_2=\frac{4}{3}$) and the vacuum ($C_2=0$) is

$$\omega(0)=\frac{1}{3\pi^2}\int d^3q\,V_C(|\boldsymbol{q}|)+\text{infrared-finite}\,, \tag{4.24}$$

in agreement with eqs. (4.4) and (4.18). Moreover, comparing eqs. (4.22) and (4.23), we see that the infrared divergence can be completely eliminated from all quantities,

unphysical as well as physical, by making the potential shift of eq. (4.3) with

$$C=-\frac{1}{(2\pi)^3}\int d^3q\, V_C(|\boldsymbol{q}|)\,, \tag{4.25a}$$

corresponding to the use of a modified momentum-space potential

$$V_C(|\boldsymbol{q}|)\rightarrow V_C(|\boldsymbol{q}|)-\delta^3(\boldsymbol{q})\int d^3q\, V_C(|\boldsymbol{q}|)\,. \tag{4.25b}$$

By this method, we can get manifestly infrared-finite analogs of eqs. (2.12a,b) for A and B, permitting a numerical computation to be carried out without the introduction of the pair wave function Ψ. The ideas just outlined will play an essential role in an extension of the numerical solution of sect. 5 to include transverse gluon exchange, since when retardation effects are included there is no analog of the pair wave function Ψ, and one must work directly with the gap equation in the form given in eq. (3.12).

5. Numerical solution for a pure confining potential

To verify the infrared-finiteness properties discussed in the preceding section, we have solved the Coulomb gauge pairing model numerically for the case of a pure confining potential,

$$V(r)=\kappa r\,,\qquad V_C(|\boldsymbol{q}|)=\frac{2\kappa}{(\boldsymbol{q}^2)^2}\,. \tag{5.1a}$$

Eq. (5.1a) of course does not correctly represent the high-momentum behavior of V_C, which is dominated by single-Coulomb gluon exchange. However, at high momenta the transverse (V_T) term in eq. (3.4) is expected to be as important as the Coulomb (V_C) term, and both should be included in any realistic analysis of the high-momentum regime and of renormalization effects. We hope to extend our analysis later on to include high-momentum components of the Bethe–Salpeter kernel.

Since the potential of eq. (5.1a) does not lead to ultraviolet divergences, we work with the unrenormalized gap equation of eq. (2.6b), and with the corresponding equation for the pion vertex g given in eq. (4.17). Introducing the angular integration kernels

$$\begin{aligned}
I_{(2)}(p,q)&=\int_{-1}^{1}d(\hat{p}\cdot\hat{q})\,V_C(|\boldsymbol{p}-\boldsymbol{q}|)=\frac{4\kappa}{(p^2-q^2)^2}\,,\\
I_{(1)}(p,q)&=\int_{-1}^{1}d(\hat{p}\cdot\hat{q})\,\hat{p}\cdot\hat{q}\,V_C(|\boldsymbol{p}-\boldsymbol{q}|)\\
&=\frac{p^2+q^2}{2pq}I_{(2)}(p,q)-\frac{\kappa}{p^2q^2}\log\left|\frac{p+q}{p-q}\right|\,.
\end{aligned} \tag{5.1b}$$

Eqs. (2.6b) and (4.17) reduce to one-dimensional integral equations

$$p\Psi(p)=\frac{2}{3\pi}\int_0^\infty \mathrm{d}q\, q^2\frac{I_{(2)}(p,q)\Psi(q)[1-\Psi^2(p)]-I_{(1)}(p,q)\Psi(p)[1-\Psi^2(q)]}{1+\Psi^2(q)}, \tag{5.2a}$$

$$g(p)p\cos 2\vartheta(p)=\sin 2\vartheta(p)+\frac{2}{3\pi}\int_0^\infty \mathrm{d}q\, q^2[\sin 2\vartheta(p)\sin 2\vartheta(q)I_{(2)}(p,q)$$
$$+\cos 2\vartheta(p)\cos 2\vartheta(q)I_{(1)}(p,q)][g(q)-g(p)]. \tag{5.2b}$$

To solve these equations numerically, we use relaxation methods appropriate for a nonlinear problem, as discussed in a recent pedagogical review by Adler and Piran [14]. The equations are reduced to discrete form on a mesh containing node and half-node points, with p values on the node mesh and (to avoid singularities of $I_{(1,2)}$ at $p=q$) with q values on the half-node mesh. As a function of a particular node value $\psi=\Psi(p_j)$, eq. (5.2a) clearly takes the form

$$\psi C_1+\frac{\psi(1-\psi^2)}{1+\psi^2}C_2+(1-\psi^2)C_3=0, \tag{5.3}$$

with $C_{1,2,3}$ functions of the node values $\Psi(p_i)$, $i\neq j$. The basic iteration used for ψ begins with two Newton iterations of eq. (5.3), starting from the old value of ψ as the initial guess. The Newton iterations yield an improved value of ψ which is then used to update the old value through use of the over-relaxed Gauss–Seidel algorithm. (The procedure is the same as that used to solve the classical abelian Higgs model in ref. [14].) Once Ψ has been iterated to convergence, $\sin 2\vartheta(p_j)$ and $\cos 2\vartheta(p_j)$ are computed from eq. (2.10) and are substituted into the discrete form of eq. (5.2b). This yields a linear discrete problem for the node values $g(p_j)$, which is again solved by iteration using the over-relaxed Gauss–Seidel algorithm.

The numerical solution for $\Psi(p)$ is plotted in fig. 1, using momentum units in which $1=\kappa^{1/2}\approx 350$ MeV. For small momentum, Ψ decreases linearly as

$$\Psi(p)\approx 1-\alpha p, \qquad \alpha\approx 5. \tag{5.4}$$

The value of the small-momentum slope of Ψ can be related to an effective quasi-particle mass m^*, as follows. Consider first the propagator for a free massive Dirac particle,

$$\frac{1}{\gamma_0 p_0-\boldsymbol{\gamma}\cdot\boldsymbol{p}-m}=\frac{\gamma_0 p_0-\boldsymbol{\gamma}\cdot\boldsymbol{p}+m}{[p_0-\sqrt{\boldsymbol{p}^2+m^2}][p_0+\sqrt{\boldsymbol{p}^2+m^2}]}$$
$$=\frac{R(\boldsymbol{p})}{p_0-\sqrt{\boldsymbol{p}^2+m^2}}+\text{analytic at } p_0=\sqrt{\boldsymbol{p}^2+m^2}, \tag{5.5a}$$

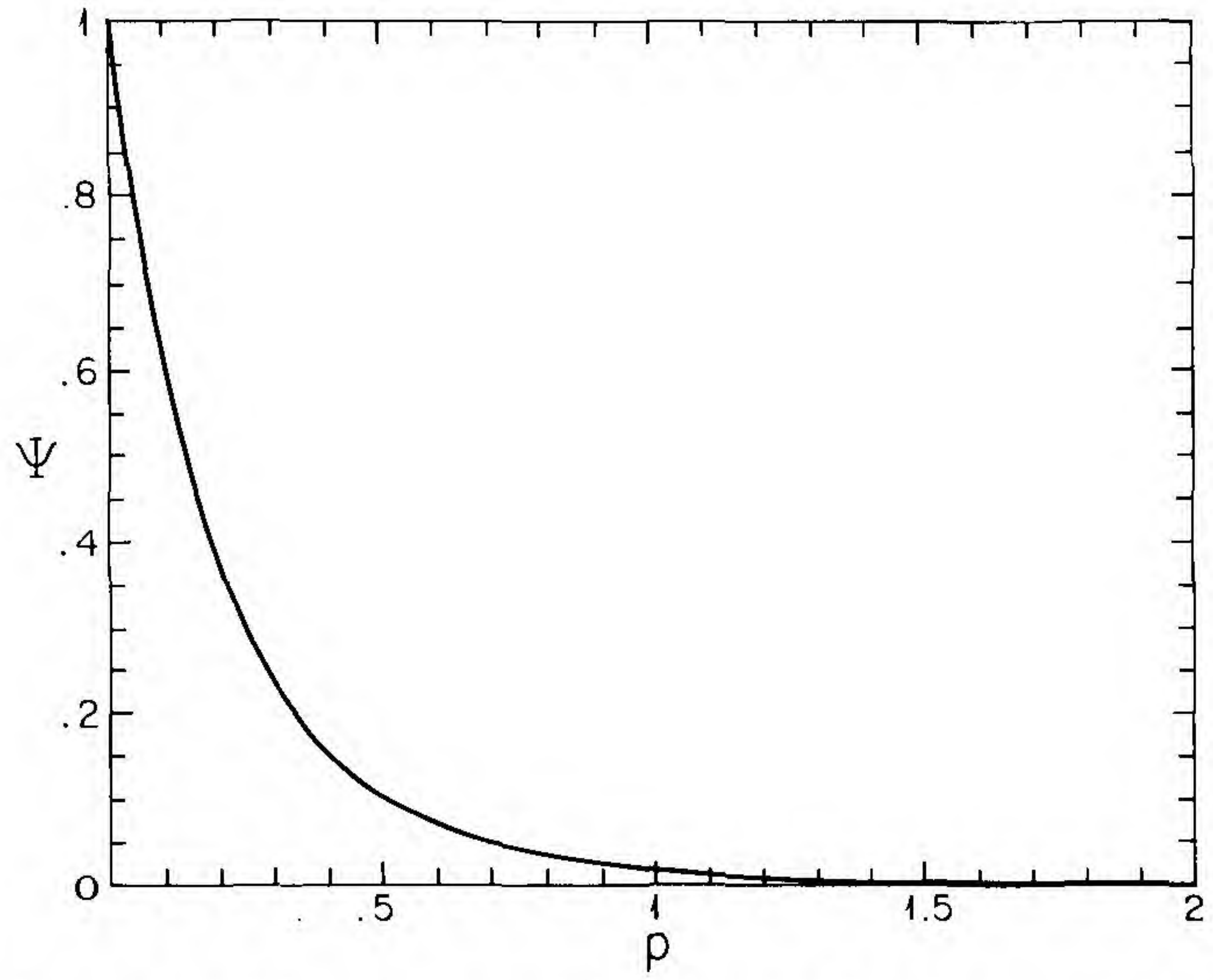


Fig. 1. The gap function $\Psi(p)$ for the pure confining potential of eq. (5.1), plotted versus momentum in units with $\kappa^{1/2}=1$. The numerical computation employs a jacobian transformation $p=v^2/(1-v)$, with a mesh of 200 points uniformly distributed in the interval $0\leq v\leq 1$.

with the residue $R(p)$ at the positive frequency pole given by

$$R(p)=\frac{\gamma_0\sqrt{p^2+m^2}-\gamma\cdot p+m}{2\sqrt{p^2+m^2}}=\tfrac{1}{2}(1+\gamma_0)-\frac{\gamma\cdot p}{2m}+\mathrm{O}(p^2)\,. \tag{5.5b}$$

Let us compare this with the propagator in the pairing model as given by eq. (4.18),

$$\frac{1}{\gamma_0 p_0-\gamma\cdot p-\Sigma}=\frac{R_+(p)}{p_0-\omega(p)}+\text{analytic at } p_0=\omega(p)\,, \tag{5.6a}$$

with the residue at the positive frequency pole now given by

$$\begin{aligned}R_+(p)&=\frac{1}{2}\left[\frac{2\Psi(p)}{1+\Psi^2(p)}+\gamma_0\right]-\tfrac{1}{2}\gamma\cdot p\frac{1}{p}\frac{1-\Psi^2(p)}{1+\Psi^2(p)}\\&=\tfrac{1}{2}(1+\gamma_0)-\tfrac{1}{2}\gamma\cdot p\alpha+\mathrm{O}(p^2)\,.\end{aligned} \tag{5.6b}$$

The small-momentum expansion of eq. (5.6b) clearly has the same form as that in eq. (5.5b), with an effective mass m^* given by

$$m^*=\frac{1}{\alpha}\approx\tfrac{1}{5}\times 350\text{ MeV}\approx 70\text{ MeV}\,. \tag{5.7}$$

From the solution for $\Psi(p)$, we can also evaluate the quark condensate $\langle\bar{u}u\rangle$ by evaluating the integral in eq. (4.9), with the result

$$\langle\bar{u}u\rangle\approx(-95\text{ MeV})^3\,. \tag{5.8}$$

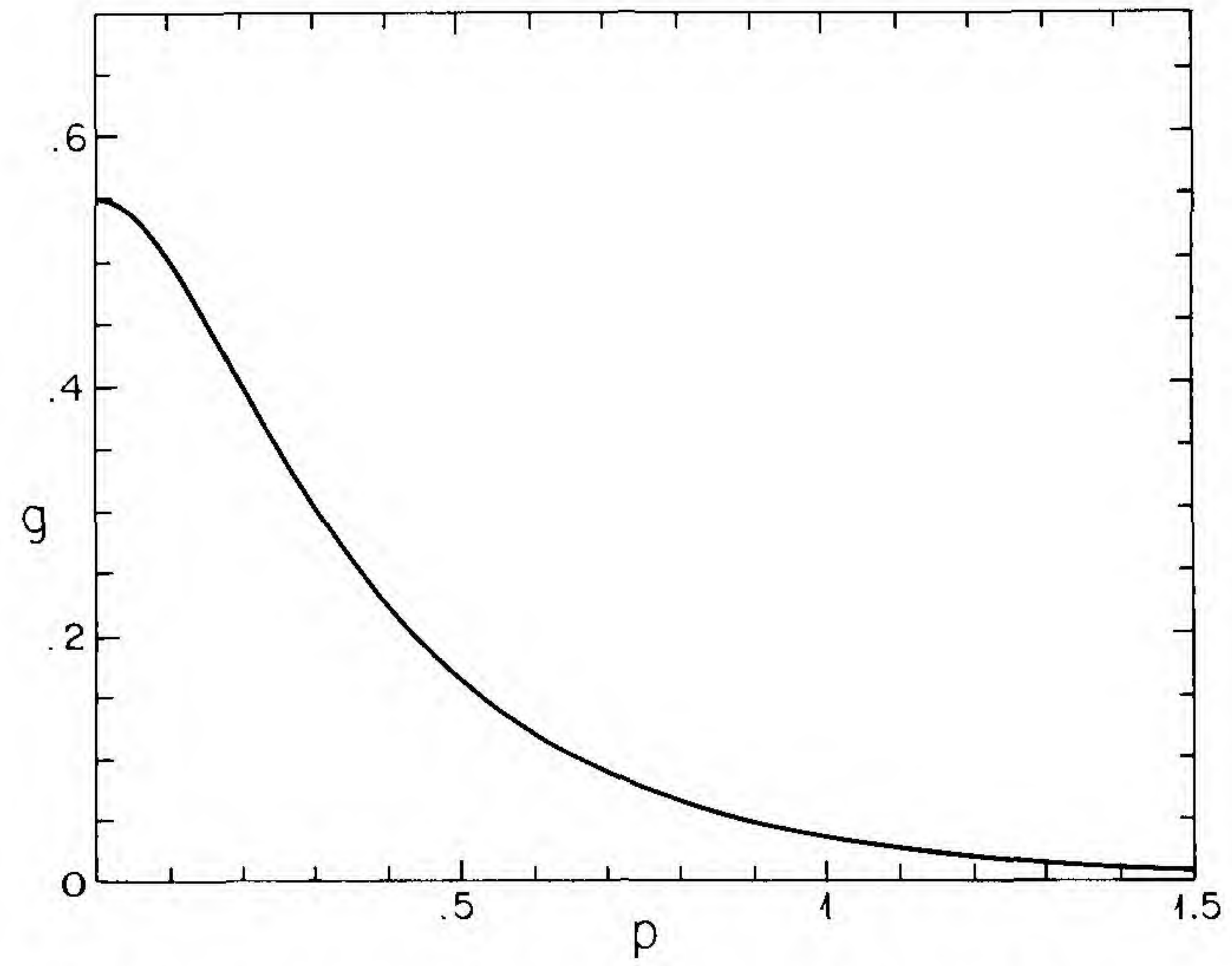


Fig. 2. The pion vertex function $g(p)$ corresponding to the gap function of fig. 1.

The numerical solution for $g(p)$ is plotted in fig. 2, again using momentum units in which $\kappa^{1/2}=1$. Using this solution, we can calculate the pion decay constant f_π by evaluating the integral in eq. (4.13), with the result

$$f_\pi \approx 11\ \mathrm{MeV}. \tag{5.9}$$

Experimentally, the constituent quark mass m^*, the quark condensate $\langle \bar{u}u \rangle$ and the pion decay constant f_π have the values

$$\begin{aligned} m^*_{\mathrm{expt}} &= 300\ \mathrm{MeV}, \\ \langle \bar{u}u \rangle_{\mathrm{expt}} &\approx (-230\ \mathrm{MeV})^3, \\ f_{\pi\,\mathrm{expt}} &\approx 95\ \mathrm{MeV}. \end{aligned} \tag{5.10}$$

Thus, the pairing model with a pure confining potential gives values for these which are consistently too small. Since the high-momentum part of V_C has the same sign as the confining term (cf. eq. (3.7)), its inclusion should have the same qualitative effect as increasing the string tension $\kappa^{1/2}$, which in the model of eq. (5.1) increases m^*, $\langle \bar{u}u \rangle$ and f_π. Hence, improved agreement with experiment is likely to result when high-momentum components of the Bethe–Salpeter kernel are included in the calculation.

Finally, in fig. 3 we plot the excitation energy $\omega(p)-\omega(0)$, as calculated from the numerical solution using eq. (4.6b). Both the numerical work and analytic estimates obtained from eq. (4.6) show that the excitation energy vanishes at small momenta as p^2 (up to a possible factor of $|\log p^2|$), and at large momenta approaches

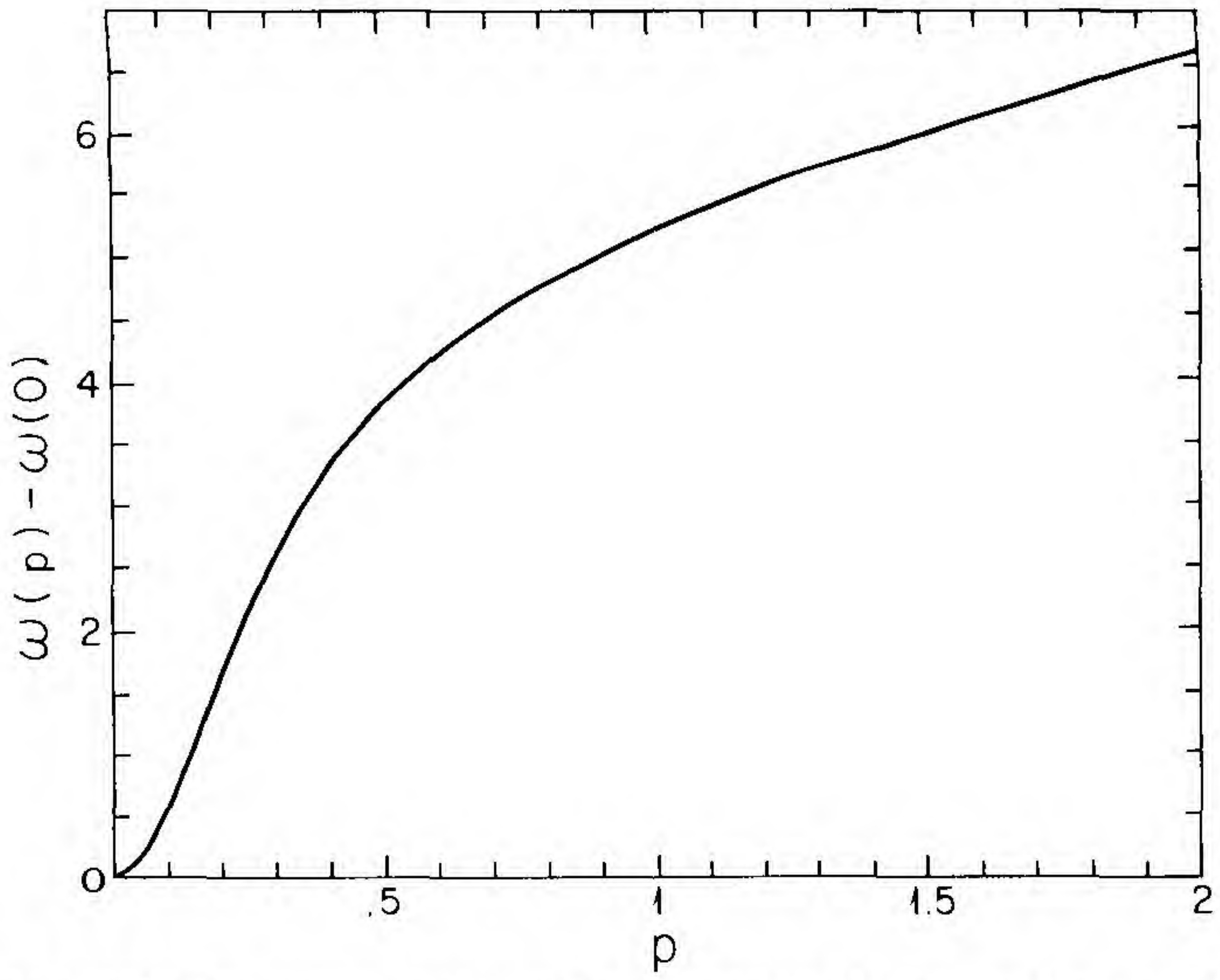


Fig. 3. The excitation energy $\omega(p)-\omega(0)$ corresponding to the gap function of fig. 1.

the excitation energy for a massless free particle to within a finite additive constant,

$$\omega(p)-\omega(0) \underset{p\to\infty}{\longrightarrow} p+\frac{1}{3\pi^2}\int d^3q\, V_C(|\boldsymbol{q}|)[\sin 2\vartheta(q)-1]+O\left(\frac{1}{p}\right). \quad (5.11)$$

Hence although the Coulomb gauge pairing model is not Lorentz-covariant, the high-momentum behavior of the excitation energy in this model nonetheless joins smoothly onto a relativistic dispersion law.

6. Conclusions

By use of the renormalized vector and axial-vector vertex equations and Ward identities we have constructed the correctly renormalized gap equation for chiral symmetry breaking in Coulomb gauge QCD. Making the ladder approximation to the Bethe–Salpeter kernel relates the chiral symmetry breaking condensate, and the other chiral parameters, to the static quark potential. This allows a phenomenological potential to be used in the study of chiral symmetry breaking. We have demonstrated the infrared-finiteness of all physical parameters, and have elucidated the structure of the infrared divergences. Neglecting the effect of transverse gluons we have numerically evaluated $\langle\bar{u}u\rangle$, f_π and the effective quark mass for a pure confining potential, $V_C(q)\sim 1/q^4$. The values we obtain are rather low compared with those deduced from experiment. However, inclusion of the high momentum tail should give better results.

Inclusion of transverse gluons and the Coulomb tail in our gap equation will permit a more detailed investigation of the relationship between chiral symmetry

breaking and confinement in QCD. Further, extending our analysis to finite temperatures will allow a study of the relationship between the deconfinement transition and the chiral symmetry restoration transition. Work on these questions is in progress.

Appendix A

BOGOLIUBOV-VALATIN TRANSFORMATION

The Bogoliubov–Valatin transformation for the pairing model relates the operators b, d which annihilate $|0\rangle$ to a new basis set B, D which annihilate $|\Psi\rangle$,

$$B^{(a)}(\boldsymbol{p}, s) = \frac{1}{\sqrt{1+\Psi^2(p)}}[b^{(a)}(\boldsymbol{p}, s) + s\Psi(p)d^{(a)\dagger}(-\boldsymbol{p}, s)],$$

$$D^{(a)}(\boldsymbol{p}, s) = \frac{1}{\sqrt{1+\Psi^2(p)}}[d^{(a)}(\boldsymbol{p}, s) - s\Psi(p)b^{(a)\dagger}(-\boldsymbol{p}, s)]. \tag{A.1}$$

The quark field can be reexpressed in terms of this basis, giving

$$q^{(a)}(\boldsymbol{x}, 0) = \int \frac{d^3p}{(2\pi)^{3/2}} \sum_s [B^{(a)}(\boldsymbol{p}, s)M_1(\boldsymbol{p}, s)\,e^{i\boldsymbol{p}\cdot\boldsymbol{x}} + D^{(a)\dagger}(\boldsymbol{p}, s)M_2(\boldsymbol{p}, s)\,e^{-i\boldsymbol{p}\cdot\boldsymbol{x}}],$$

$$M_1(\boldsymbol{p}, s) = \frac{1}{\sqrt{1+\Psi^2(p)}}(1+\gamma_0\Psi)U(\boldsymbol{p}, s),$$

$$M_2(\boldsymbol{p}, s) = \frac{1}{\sqrt{1+\Psi^2(p)}}(1-\gamma_0\Psi)V(\boldsymbol{p}, s), \tag{A.2}$$

with U, V helicity spinors which satisfy

$$\gamma_0 sV(-\boldsymbol{p}, s) = U(\boldsymbol{p}, s),$$

$$\sum_s U(\boldsymbol{p}, s)U^\dagger(\boldsymbol{p}, s) = \sum_s V(\boldsymbol{p}, s)V^\dagger(\boldsymbol{p}, s) = \tfrac{1}{2}(1-\boldsymbol{\gamma}\cdot\hat{\boldsymbol{p}}\gamma_0). \tag{A.3}$$

It is often convenient to characterize the Bogoliubov–Valatin transformation by a rotation angle $\vartheta(p)$ defined by

$$\sin\vartheta(p) = \frac{\Psi(p)}{\sqrt{1+\Psi^2(p)}}, \qquad \cos\vartheta(p) = \frac{1}{\sqrt{1+\Psi^2(p)}}. \tag{A.4}$$

References

[1] S.J. Brodsky, Perturbative quantum chromodynamics, in Proc. 1979 Summer Institute on Particle Phys., SLAC report no. 224, 1980
[2] L.H. Karsten and J. Smit, Nucl. Phys. B183 (1981) 103
[3] G. 't Hooft, in Recent developments in gauge theories, ed. G. 't Hooft, et al. (Plenum, New York, 1980);
M.E. Peskin, Chiral symmetry and chiral symmetry breaking, chap. 6, SLAC-PUB-3021, 1982, to appear in Proc. 1981 Les Houches Summer School

[4] K.D. Lane, Phys. Rev. D10 (1974) 1353;
M.E. Peskin, ref. [3], chap. 3
[5] J.R. Finger and J.E. Mandula, Nucl. Phys. B199 (1982) 168;
J.R. Finger, D. Horn and J. Mandula, Phys. Rev. D20 (1979) 3253
[6] E. Eichten, K. Gottfried, T. Kinoshita, K.D. Lane and T.M. Yan, Phys. Rev. D21 (1980) 203;
E. Eichten and F. Feinberg, Phys. Rev. D23 (1981) 2724
[7] A. Amer, A. Le Yaouanc, L. Oliver, O. Pène and J-C. Raynal, Phys. Rev. Lett. 50 (1983) 87; Phys. Rev. D28 (1983) 1530;
A. Le Yaouanc, L. Oliver, O. Pène and J-C. Raynal, Phys. Rev. D29 (1984) 1233
[8] J. Govaerts, J.E. Mandula and J. Weyers, Phys. Lett. 130B (1983) 427
[9] J.R. Schrieffer, Theory of superconductivity (Benjamin, New York, 1964) sect. 8.5
[10] J.D. Bjorken and S.D. Drell, Relativistic quantum fields (McGraw-Hill, New York, 1965) sects. 19.3 and 19.8
[11] S.L. Adler, Phys. Rev. 177 (1969)2426
[12] S.L. Adler, Quark statics, in Workshop on non-perturbative quantum chromodynamics, ed. K.A. Milton and M.A. Samuel, Progress in Physics, vol. 8 (Birkhaûser, Boston, 1983)
[13] A. Le Yaouanc, L. Oliver, O. Pène and J-C. Raynal, Phys. Lett. 134B (1984) 249
[14] S.L. Adler and T. Piran, Rev. Mod. Phys. 56 (1984) 1

12 Progress of Theoretical Physics Supplement No. 86, 1986

Gap Equation Models for Chiral Symmetry Breaking

Stephen L. ADLER

The Institute for Advanced Study, Princeton, New Jersey 08540

(Received October 26, 1985)

We give a critical discussion of gap equation models for chiral symmetry breaking, and formulate an extended Coulomb gauge model which includes one gluon exchange (but does not resolve the problems pointed out in our critique).

§ 1. Introduction

Quantum chromodynamics (QCD) is now widely accepted as the theory of the strong interactions, and gives a concrete realization of the prophetic idea of Nambu and Jona-Lasinio[1] that the pion is an almost massless fermion-antifermion bound state. A number of recent pairing model calculations,[2),3)] based on approximations to QCD, have shown that chiral symmetry breaking, and the accompanying generation of a Nambu-Goldstone pion, necessarily occur when the instantaneous potential has a confining piece. However, the quantitative results for a phenomenological pure confining potential are not good, leading to values for the quark condensate $[-\langle \bar{u}u\rangle]^{1/3}$ and the pion decay constant f_π which are too small by factors of 2–5. We show below that assuming a dominant confining potential is in fact problematic, since the chiral breaking model requires a Lorentz vector instantaneous potential, whereas recent quark spectroscopic data[4] shows that the confining potential is predominantly Lorentz scalar. One natural way to try to improve the model is by including the leading high-momentum components of the quark-antiquark potential, which arise from one gluon exchange and are Lorentz vector in structure. In this article we work out (but do not attempt to solve) the equations for this extension of the model. In §2 we review the general gap equation formalism, critically discuss confining-potential models, and show that including the instantaneous Coulomb potential without including transverse gluon exchange is inconsistent. We formulate an extended model including both Coulomb and transverse gluon exchange, and give the integral equations for this model in §3. In §4 we conclude with a brief discussion.

§ 2. General formalism and discussion of gap equation potential models

We sketch a method[3] for obtaining the gap equation, including renormalization counter-terms, in a general gauge in which the vertex renormalization is not finite. The basic idea[5] is to use the Ward identities to derive the renormalized gap equation from suitable approximations to the vertex parts. Specifically, we start from the renormalized Dyson equation for the vector and axial-vector vertices, making the ladder approximation

that the Bethe-Salpeter kernel depends only on the momentum transfer. Consistent with this approximation, we exclude quark annihilation graphs, so that there are no anomalies, and neglect terms arising from the non-commutativity of color matrices on the quark lines, so that we can use the Ward identities of QED and can omit all color indices. With these simplifications, the renormalized vector and axial-vector vertex parts satisfy the integral equations

$$\begin{aligned}\tilde{\Gamma}_\mu(p', p)_{\delta\gamma} &= (Z\gamma_\mu)_{\delta\gamma} + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}'_F(p'+q)\tilde{\Gamma}_\mu(p'+q, p+q)i\tilde{S}'_F(p+q)]_{\beta\alpha} \\ &\quad \times \bar{K}_{\alpha\beta,\gamma\delta}(p+q, p'+q, q), \\ \tilde{\Gamma}_{\mu5}(p', p)_{\delta\gamma} &= (Z\gamma_\mu\gamma_5)_{\delta\gamma} + \int \frac{d^4q}{(2\pi)^4}[i\tilde{S}'_F(p'+q)\tilde{\Gamma}_{\mu5}(p'+q, p+q)i\tilde{S}'_F(p+q)]_{\beta\alpha} \\ &\quad \times \bar{K}_{\alpha\beta,\gamma\delta}(p+q, p'+q, q) \qquad (1)\end{aligned}$$

with $\tilde{S}'_F$ the full renormalized quark propagator and with $\bar{K}_{\alpha\beta,\gamma\delta}$ the renormalized quark-antiquark Bethe-Salpeter kernel. Since anomalies have been excluded, the vector and axial-vector vertex parts have the same renormalization constants, and satisfy the Ward identities

$$\begin{aligned}(p'-p)^\mu \tilde{\Gamma}_\mu(p', p) &= \tilde{S}'_F{}^{-1}(p') - \tilde{S}'_F{}^{-1}(p), \\ (p'-p)^\mu \tilde{\Gamma}_{\mu5}(p', p) &= \gamma_5 \tilde{S}'_F{}^{-1}(p) + \tilde{S}'_F{}^{-1}(p')\gamma_5 . \qquad (2)\end{aligned}$$

We now make the fundamental approximation of assuming that the Bethe-Salpeter kernel $\bar{K}_{\alpha\beta,\gamma\delta}(p+q, p'+q, q)$ is a function only of the momentum transfer q,

$$\bar{K}_{\alpha\beta,\gamma\delta}(p+q, p'+q, q) \approx \bar{k}_{\alpha\beta,\gamma\delta}(q), \qquad (3a)$$

and has *Lorentz vector* couplings on the quark and antiquark lines so that there is no *explicit* breaking of chiral symmetry,

$$\begin{aligned}\bar{k}_{\alpha\beta,\gamma'\delta}(q)(\gamma_5)_{\gamma'\gamma} &= -(\gamma_5)_{\alpha\alpha'}\bar{k}_{\alpha'\beta,\gamma\delta}(q), \\ \bar{k}_{\alpha\beta,\gamma\delta'}(q)(\gamma_5)_{\delta'\delta} &= -(\gamma_5)_{\beta\beta'}\bar{k}_{\alpha\beta',\gamma\delta}(q). \qquad (3b)\end{aligned}$$

Substituting Eq. (3) into Eq. (1), further substituting Eq. (1) into the Ward identities of Eq. (2) and using the Ward identities a second time to rearrange the integrands, we find that Eqs. (1) ~ (3) imply that the renormalized quark propagator must satisfy the integral equation

$$\tilde{S}'_F{}^{-1}(p) = (Z\gamma_\mu)p^\mu + \int \frac{d^4q}{(2\pi)^4}\tilde{S}'_F(p+q)_{\beta\alpha}\bar{k}_{\alpha\beta,\cdots}(q) . \qquad (4a)$$

Introducing the self-energy Σ by writing

$$\tilde{S}'_F{}^{-1}(p) = \gamma_\mu p^\mu - \Sigma(p), \qquad (4b)$$

and writing out all Dirac indices explicitly, we see that Σ must satisfy the integral equation

$$\Sigma(p)_{\delta\gamma} = [\gamma_\mu - (Z\gamma_\mu)]_{\delta\gamma}p^\mu - \int \frac{d^4q}{(2\pi)^4}\tilde{S}'_F(p+q)_{\beta\alpha}\bar{k}_{\alpha\beta,\gamma\delta}(q). \qquad (5)$$

This is the renormalized gap equation corresponding to the ladder approximation to the Bethe-Salpeter kernel.

As discussed in Ref. 3), if one assumes that $\bar{k}_{\alpha\beta,\gamma\delta}(q)$ contains only an instantaneous potential

$$\bar{k}^{I}_{\alpha\beta,\gamma\delta}(\boldsymbol{q}) = -4\pi i(\gamma_0)_{\delta\beta}(\gamma_0)_{\alpha\gamma}\frac{4}{3}V_C(|\boldsymbol{q}|), \tag{6}$$

then Eq. (5) is readily reduced[2),3)] to the gap equation for a non-relativistic pairing model. If one takes V_C to be the pure confining potential

$$V_C = \frac{3\sigma/2}{(\boldsymbol{q}^2)^2}, \quad \sigma = \text{string tension}, \tag{7}$$

the gap equation has a non-trivial solution, indicating a spontaneous breakdown of chiral symmetry; but as noted above, use of the phenomenological value $\sigma^{1/2} = 400\text{MeV}$ gives poor results for the parameters characterizing chiral symmetry breaking. However, there is a serious inconsistency in assuming the gap equation to be dominated by a phenomenological confining potential. The derivation of the gap equation *requires* the instantaneous potential to have the Lorentz vector couplings of Eq. (6), since a Lorentz scalar instantaneous potential

$$\bar{k}^{I}_{\alpha\beta,\gamma\delta}(\boldsymbol{q}) = -4\pi i(1)_{\delta\beta}(1)_{\alpha\gamma}\frac{4}{3}V_S(|\boldsymbol{q}|) \tag{8}$$

would manifestly break chiral symmetry. On the other hand, experimental data on heavy quark spectroscopy[4)] shows that the confining potential in heavy quark systems is predominantly Lorentz scalar. Moreover, theoretical arguments[6)] based on the behavior of electric flux tubes also suggest that in the phenomenological confining potential

$$V(r) = \sigma r - \frac{4}{3}\frac{\alpha}{r}, \tag{9}$$

only the Coulombic piece $-(4/3)(\alpha/r)$ is Lorentz vector, while the confining piece σr is Lorentz scalar. We conclude that there is a puzzle here, and quite likely an indication that the approximations leading to the gap equation are not valid for the confining part of the potential.

To get a possibly more realistic model, within the framework of the approximations leading to Eq. (5), one should clearly include the Coulombic piece (which *is* Lorentz vector) in V_C, by replacing Eq. (7) by

$$V_C = \frac{3\sigma/2}{(\boldsymbol{q}^2)^2} + \frac{\alpha}{\boldsymbol{q}^2}. \tag{10}$$

However, as shown in Ref. 3), Eq. (10) leads to the non-covariant counter-term structure

$$\text{“}Z\gamma_\mu\text{''} = \begin{cases} \gamma_0, & \mu = 0, \\ Z\gamma_j, & \mu = j = 1,2,3, \quad Z = \text{log divergent}, \end{cases} \tag{11}$$

whereas the 0 and j components of the vertex are expected to have the same logarithmic divergence even in a non-covariant gauge. Thus, Lorentz covariance of the counter-term requires inclusion of transverse gluon exchange along with the Coulomb gluon term, by

writing

$$\bar{k}_{\alpha\beta,\gamma\delta}=\bar{k}^{I}_{\alpha\beta,\gamma\delta}(\boldsymbol{q})+\bar{k}^{T}_{\alpha\beta,\gamma\delta}(q)\,,$$

$$\bar{k}^{T}_{\alpha\beta,\gamma\delta}(q)=-4\pi i(\gamma_j)_{\delta\beta}(\gamma_k)_{\alpha\gamma}\left(\delta_{jk}-\frac{q_jq_k}{\boldsymbol{q}^2}\right)\frac{4}{3}V_T(q),$$

$$V_T(q)=\frac{\alpha}{q_0{}^2-\boldsymbol{q}^2}\,. \tag{12}$$

A simple calculation shows that when Eqs. (12), (10) and (6) are substituted into the gap equation of Eq. (5), all divergences are removed with a covariant counter-term $Z\gamma_\mu$, with

$$Z=1-\frac{1}{4\pi^2}\int d^3q\frac{2}{3}\frac{\alpha}{q^3}+\text{finite}\,. \tag{13}$$

One further point must be addressed to formulate a model with Coulomb gluon exchange included. The coordinate-space potential of Eq. (9) is undetermined up to an additive constant, and this corresponds to the freedom to add to the Fourier transform V_C a multiple of $\delta^3(\boldsymbol{q})$. We will choose this multiple so that $1/(\boldsymbol{q}^2)^2$ is replaced by

$$[1/(\boldsymbol{q}^2)^2]_{\text{REG}}=1/(\boldsymbol{q}^2)^2-\delta^3(q)\int d^3q'/(\boldsymbol{q}'^2)^2\,, \tag{14a}$$

which by construction satisfies

$$\int d^3q[1/(\boldsymbol{q}^2)^2]_{\text{REG}}=0\,. \tag{14b}$$

The infrared subtraction in Eq. (14) guarantees that the confining piece represents a linear potential which vanishes at $r=0$; without the subtraction, the Fourier transform of Eq. (10) gives a linear potential which equals a linearly divergent constant at $r=0$. Use of Eq. (14a) yields a gap equation which is manifestly infrared finite, even when retarded potentials are included. [When retarded potentials are neglected, Eq. (14a) makes the A and B functions of Ref. 3) infrared finite, whereas if one uses Eq. (7), only the gap function Ψ(which has no analog in the retarded case) is infrared-finite.] Of course, given the problems with the Lorentz structure of the confining potential, one can question whether it should be included at all! If it is omitted, and only the one-gluon exchange potential is retained, Coulomb gauge is a poor choice: For the pure gluon-exchange problem, it has long been known that the retarded integral equations take a much simpler form in a covariant gauge,[7] especially in Landau gauge where the vertex renormalization Z is finite.

§ 3. Equations of the retarded model

To summarize, the retarded model has a gap equation given by Eq. (5), with

$$\bar{k}_{\alpha\beta,\gamma\delta}=-4\pi i(\gamma_0)_{\delta\beta}(\gamma_0)_{\alpha\gamma}\frac{4}{3}\left\{\frac{3}{2}\sigma\left[\frac{1}{(\boldsymbol{q}^2)^2}-\delta^3(q)\int\frac{d^3q'}{(\boldsymbol{q}'^2)^2}\right]+\frac{\alpha}{\boldsymbol{q}^2}\right\}$$

$$-4\pi i(\gamma_j)_{\delta\beta}(\gamma_k)_{\alpha\gamma}\left(\delta_{jk}-\frac{q_jq_k}{\boldsymbol{q}^2}\right)\frac{4}{3}\frac{\alpha}{q_0{}^2-\boldsymbol{q}^2}\,. \tag{15}$$

We assume for Σ the general Ansatz

$$\Sigma = pA(p, p_0) + \boldsymbol{\gamma}\cdot\boldsymbol{p}B(p, p_0) - \gamma_0 p_0 D(p, p_0),$$
$$p = |\boldsymbol{p}|\,. \tag{16}$$

After rotation to the Euclidean section where $p_0 = i\omega$, Σ has the form

$$\Sigma_E = pA[p, \omega] + \boldsymbol{\gamma}\cdot\boldsymbol{p}B[p, \omega] - i\gamma_0\omega D[p, \omega],$$
$$A[p, \omega] = A(p, i\omega), \text{ etc.,} \tag{17}$$

and the quark condensate is given by

$$\langle \bar{u}u \rangle = 3\int \frac{d^3p\,d\omega}{(2\pi)^4}\mathrm{Tr}[\gamma_0 i\omega - \boldsymbol{\gamma}\cdot\boldsymbol{p} - \Sigma_E]^{-1}\,. \tag{18}$$

After some algebra, the gap equation can be reduced to the following coupled integral equations for A, B and D on the Euclidean section,

$$pA[p, \omega] = \frac{1}{4\pi^3}\int_0^\infty d\zeta \int d^3q \left\{ \frac{3\sigma/2}{[(\boldsymbol{q}-\boldsymbol{p})^2]^2}\left[\frac{2qA[q, \zeta]}{d[q, \zeta]} - \frac{2pA[p, \zeta]}{d[p, \zeta]}\right]\right.$$
$$+ \frac{2qA[q, \zeta]}{d[q, \zeta]}\left[\frac{4}{3}\frac{\alpha}{(\boldsymbol{q}-\boldsymbol{p})^2} + \frac{4}{3}\frac{\alpha}{(\zeta-\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2}\right.$$
$$\left.\left. + \frac{4}{3}\frac{\alpha}{(\zeta+\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2}\right]\right\},$$

$$pB[p, \omega] = p(Z-1) + \frac{1}{4\pi^3}\int_0^\infty d\zeta \int d^3q \left\{ \frac{3\sigma/2}{[(\boldsymbol{q}-\boldsymbol{p})^2]^2}\left[\frac{\mu 2qC[q, \zeta]}{d[q, \zeta]} - \frac{2pC[p, \zeta]}{d[p, \zeta]}\right]\right.$$
$$+ \frac{2qC[q, \zeta]}{d[q, \zeta]}\left[\frac{4}{3}\frac{\alpha\mu}{(\boldsymbol{q}-\boldsymbol{p})^2} - \frac{4}{3}\alpha\frac{qp(1+\mu^2) - (p^2+q^2)\mu}{p^2 + q^2 - 2pq\mu}\right.$$
$$\left.\left.\times\left(\frac{1}{(\zeta-\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2} + \frac{1}{(\zeta+\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2}\right)\right]\right\},$$

$$\omega D[p, \omega] = \omega(Z-1) + \frac{1}{4\pi^3}\int_0^\infty d\zeta \int d^3q \frac{2\zeta E[q, \zeta]}{d[q, \zeta]}$$
$$\times \frac{4}{3}\alpha\left[\frac{1}{(\zeta-\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2} - \frac{1}{(\zeta+\omega)^2 + (\boldsymbol{q}-\boldsymbol{p})^2}\right] \tag{19a}$$

with

$$E[p, \omega] = 1 + D[p, \omega], \quad C[p, \omega] = 1 + B[p, \omega],$$
$$d[p, \omega] \equiv \omega^2 E^2[p, \omega] + p^2\{C^2[p, \omega] + A^2[p, \omega]\},$$
$$\mu = \boldsymbol{p}\cdot\boldsymbol{q}/(pq). \tag{19b}$$

In writing Eq. (19), we have made use of the following reflection symmetries,

$$A, B, C, D, E \text{ are invariant under } \omega \to -\omega\,. \tag{20}$$

§ 4. Discussion

Getting a quantitative realization of the Nambu-Jona-Lasinio model within QCD has turned out to be more difficult than one would have hoped. Coulomb gauge gap equation models based on an instantaneous confining potential give chiral symmetry breaking, but are at variance with the observed Lorentz structure of the phenomenological confining potential. Extending the model to include Coulomb gluon exchange does not cure the Lorentz structure problem and requires the inclusion of transverse gluons as well; this leads (after angular averaging) to a set of coupled two-dimensional integral equations, which will be much harder to study analytically[8] and to solve numerically than the one-dimensional integral equations of the instantaneous potential model. Moreover, the retarded equations are renormalization scheme dependent, since Eq. (19) contains the *finite part* of the renormalization constant Z as a parameter. Thus comparison of the extended model with experiment will have to take the renormalization scheme dependence[9] of $\langle \bar{u}u \rangle$ into account, a complicating feature which again was absent in the instantaneous potential models. The situation is sufficiently complex that, without a resolution of the Lorentz structure problem, the motivation for a further elaboration of the Coulomb gauge gap equation model seems much diminished.

Acknowledgements

I wish to thank A.C. Davis for discussions about parts of this work and T. Dickens and A. M. Matheson for checking the derivation of Eq. (19). This work was supported by the U.S. Department of Energy under Grant No. DE-AC02-76ERO2220.

References

1) Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122** (1961), 345.
2) J. R. Finger and J. E. Mandula, Nucl. Phys. **B199** (1982), 168.
J. Govaerts, J. E. Mandula and J. Weyers, Phys. Lett. **130B** (1983), 427.
A. Amer, A. LeYaouanc, L. Oliver, O. Pène and J. -C. Raynal, Phys. Rev. Lett. **50** (1983), 87; Phys. Rev. **D28** (1983), 1530.
A. LeYaouanc, L. Oliver, O. Pène and J. -C. Raynal, Phys. Rev. **D29** (1984), 1233.
3) S. L. Adler and A. C. Davis, Nucl. Phys. **B244** (1984), 469.
4) J. L. Rosner, private communication; Rapporteur's talk to appear in *the Proceedings of the 1985 International Symposium on Lepton and Photon Interactions at High Energies.*
For the theoretical background, see R. McClary and N. Byers, Phys. Rev. **D28** (1983), 1692.
P. Moxhay and J. L. Rosner, Phys. Rev. **D28** (1983), 1132.
5) See also R. Delbourgo and M. D. Scadron, J. of Phys. **G5** (1979), 1621.
6) W. Buchmüller, Phys. Lett. **112B** (1982),479; "Quarkonium Spectroscopy," Lectures presented at the International School of Physics of Exotic Atoms, Erice, 31 March 6 April, 1984, CERN-TH. 3938/84.
7) M. E. Peskin, in *Recent Advances in Field Theory-and Statistical Mechanics*, Proceedings of Session XXXIX of the Université de Grenoble Summer School at Les Houches, ed. J. -B. Zuber and R. Stora, (North-Holland, Amsterdam, 1984), p. 217.
8) Some work has been done on this by A. M. Matheson (unpublished).
9) J. Gasser and H. Leutwyler, Phys. Rep. **87** (1982), 77, 90.

PHYSICAL REVIEW D VOLUME 23, NUMBER 12 15 JUNE 1981

Over-relaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 2 March 1981)

I formulate a successive over-relaxation (SOR) procedure for the Monte Carlo evaluation of the Euclidean partition function for multiquadratic actions (such as the Yang-Mills action with canonical gauge fixing). A convergence analysis for the quadratic-action (Abelian) case shows that as thermalization proceeds the mean nodal fields relax according to the difference equation arising from the standard SOR analysis of the associated classical Euclidean field equation. Hence, SOR should accelerate the thermalization process, just as it accelerates convergence in the numerical solution of second-order elliptic differential equations.

As has been much emphasized,[1] the Euclidean partition function is a fundamental tool for studying quantum field theories. For the case of a boson field theory[2] containing spin-0 scalar and spin-1 gauge fields, denoted collectively by ϕ, the partition function at inverse temperature β is given by the functional integral[3]

$$Z=\int d\phi_t \int_{\phi_t}^{\phi_t}[d\phi]\exp(-S)\,,\qquad S=\int_0^\beta dt\int d^3x\,\mathcal{L}_E\,. \tag{1}$$

In Eq. (1) $\mathcal{L}_E$ is the Euclidean action density, including source terms, and the path integral extends over periodic paths, with $\phi(0)=\phi(\beta)=\phi_t$. I will restrict my attention in the following discussion to the case where $\mathcal{L}_E$ is a multiquadratic form (that is, it is at most quadratic in each individual field component), and will assume that the Euclidean action S is bounded from below. This restriction excludes interacting spin-0 fields from consideration (renormalizability for scalars requires a ϕ^4 term in the action), but allows ϕ to contain any number of non-Abelian spin-1 gauge fields, since the outer-product form of the gauge-field self-interaction is easily seen to imply a multiquadratic action.[4,5] Of course, when gauge fields are present, the partition function as written in Eq. (1) is formally infinite, as a result of integrations over gauge transformations which leave the action invariant. In reducing Eq. (1) to a discrete form for Monte Carlo evaluation, there are two natural strategies for dealing with the gauge infinities. The first, introduced by Wilson[6] and extensively studied[7] over the past few years, consists of using a discrete procedure in such a way that an exact, but compact gauge-invariance group remains, which can then be safely included in the Monte Carlo integration.[7] While this approach has many interesting features, it suffers from the drawbacks that (1) it is expressed in terms of unitary-matrix link variables, and has no natural discrete analogs of the gauge potentials and gauge fields, and (2) the multiquadratic form of the action is lost. A second natural strategy, which I will pursue in this paper, is to use the Faddeev-Popov method[1] to break the gauge invariance. In particular, if one chooses the canonical gauge fixing[8]

$$\begin{aligned}
&b^1=0 \text{ in } R_4:\ -\infty<x_1,\dots,x_4<\infty\,,\\
&b^2=0 \text{ in } R_3:\ x_1=0,\,-\infty<x_2,x_3,x_4<\infty\,,\\
&b^3=0 \text{ in } R_2:\ x_1=x_2=0,\,-\infty<x_3,x_4<\infty\,,\\
&b^4=0 \text{ in } R_1:\ x_1=x_2=x_3=0,\,-\infty<x_4<\infty
\end{aligned} \tag{2}$$

for each gauge potential b^μ in ϕ, the gauge degeneracy is completely broken, with a Faddeev-Popov determinant which is constant. The functional integral can then be made discrete by taking the nodal values of the gauge potentials as the variables, and applying the standard replacement[9] of derivatives by finite differences to the action S. Denoting the set of node variables which are integrated over by $\{\phi\}=\{\phi(i),\ i=1,\dots,N\}$, this procedure yields a multiple integral of the form

$$Z=\left[\prod_{i=1}^{N}\int_{-\infty}^{\infty}d\phi(i)\right]e^{-S[\{\phi\}]}\,, \tag{3}$$

with S a multiquadratic form which is bounded from below. Thus, for any node variable $\phi(k)$, S can be decomposed as

$$S[\{\phi\}_k,\phi(k)]=A_k[\phi(k)-C_k]^2+B_k\,,\quad A_k>0 \tag{4}$$

with A_k, B_k, and C_k functions of the subset of node variables $\{\phi\}_k\equiv\{\phi(i),\ i=1,\dots,k-1,k+1,\dots,N\}$.

Since in typical applications the dimensionality N of the multiple integral is very large, the

numerical estimation of Eq. (3) requires use of the Monte Carlo method.[7,10] Starting from any initial configuration $\{\phi_0\}$, one generates a sequence of successive configurations, or Markov chain, $\{\phi_1\}, \{\phi_2\}, \ldots, \{\phi_M\}, \ldots$ by repeated application of a transition probability $W[\{\phi\} \to \{\phi'\}]$. The transition probability W is chosen so that in the limit as M becomes infinite, the configurations in the chain are distributed according to the equilibrium probability density $P_{eq}[\{\phi\}]$,

$$P_{eq}[\{\phi\}] = e^{-S[\{\phi\}]}. \tag{5}$$

Sufficient conditions[11] on W to guarantee an asymptotic equilibrium probability distribution are the normalization condition

$$\left[\prod_{i=1}^{N} \int d\phi(i)'\right] W[\{\phi\} \to \{\phi'\}] = 1 \text{ for all } \{\phi\}, \tag{6a}$$

the ergodicity condition

$$P_{eq}[\{\phi\}] > 0, \quad P_{eq}[\{\phi'\}] > 0 \Rightarrow W[\{\phi\} \to \{\phi'\}] > 0, \tag{6b}$$

and the detailed-balance condition

$$P_{eq}[\{\phi\}] W[\{\phi\} \to \{\phi'\}] = P_{eq}[\{\phi'\}] W[\{\phi'\} \to \{\phi\}]. \tag{6c}$$

In numerical work it is generally most convenient to change only a single node variable at a time. When specialized to this case, the form of W, for a step in which the node variable $\phi(k)$ is changed, is

$$W = w[\{\phi\}_k; \phi(k) \to \phi(k)'], \tag{7}$$

with w required to be ergodic and to satisfy the normalization and detailed-balance conditions

$$\int_{-\infty}^{\infty} d\phi(k)' w[\{\phi\}_k; \phi(k) \to \phi(k)'] = 1, \tag{8a}$$

$$\begin{aligned} P_{eq}[\{\phi\}_k, \phi(k)] w[\{\phi\}_k; \phi(k) \to \phi(k)'] \\ = P_{eq}[\{\phi\}_k, \phi(k)'] w[\{\phi\}_k; \phi(k)' \to \phi(k)]. \end{aligned} \tag{8b}$$

As is well known, the conditions of Eq. (8) do not fix w uniquely. The choice used in most Monte Carlo studies of gauge theories, motivated by the intuitive idea[7] of successively thermalizing the individual node variables, is

$$\begin{aligned} w[\{\phi\}_k; \phi(k) \to \phi(k)'] &= N[\{\phi\}_k]^{-1} e^{-S[\{\phi\}_k, \phi(k)']}, \\ N[\{\phi\}_k] &= \int_{-\infty}^{\infty} d\phi(k)' e^{-S[\{\phi\}_k, \phi(k)']}, \end{aligned} \tag{9}$$

which makes the distribution of new values $\phi(k)'$ completely independent of the old value $\phi(k)$ being replaced. For a multiquadratic action, where the dependence of S on $\phi(k)'$ is known explicitly from Eq. (4), the transition probability of Eq. (9) becomes

$$w[\{\phi\}_k; \phi(k) \to \phi(k)'] = (\pi A_k)^{-1/2} e^{-A_k[\phi(k)' - C_k]^2}. \tag{10}$$

This evidently corresponds to choosing a Gaussian distribution of the new kth-node value around a central value C_k, where C_k is the value of $\phi(k)$ which minimizes $S[\{\phi\}_k, \phi(k)]$.

As motivation for the generalization of Eq. (10) which I am about to discuss, let us briefly consider the problem of minimizing the discrete action functional $S[\{\phi\}]$. This can also be accomplished by an iterative procedure, the simplest form of which consists of starting from an initial configuration $\{\phi_0\}$, and then successively replacing each node value $\phi(k)$, $k = 1, \ldots, N$ by the value $C_k[\{\phi\}_k]$ which minimizes S. Since S is nonincreasing under this relaxation procedure, in the limit of an infinite number of steps the minimum of S (assuming it exists and is unique[12]) will be attained. However, it is well known[9] that the procedure just outlined is not the optimal point-iterative algorithm for minimizing S; much more rapid convergence to the minimum can be obtained by using the *successive over-relaxation* (SOR) method in which $\phi(k)'$ is given by

$$\begin{aligned} \phi(k)' &= \omega C_k + (1-\omega)\phi(k) \\ &= C_k + (1-\omega)[\phi(k) - C_k], \end{aligned} \tag{11}$$

with ω a parameter called the relaxation parameter. Convergence is guaranteed provided that S remains nonincreasing at each step of the iteration, which requires

$$\begin{aligned} 0 \le S[\{\phi\}_k, \phi(k)] - S[\{\phi\}_k, \phi(k)'] \\ = A_k[\phi(k) - \phi(k)'][\phi(k) + \phi(k)' - 2C_k] \\ = A_k[\phi(k) - \phi(k)']^2\left(\frac{2}{\omega} - 1\right), \end{aligned} \tag{12}$$

giving the restriction

$$0 < \omega < 2. \tag{13}$$

When $\omega = 1$, Eq. (11) reduces to $\phi(k)' = C_k$, corresponding to the simple minimization procedure in which the new value $\phi(k)'$ is independent of the old value $\phi(k)$. When $\omega \ne 1$, the new value $\phi(k)'$ clearly retains a memory of the old value $\phi(k)$. In practice, optimum convergence is obtained by doing several iterations with $\omega = 1$, and then doing many iterations with a value $\omega = \omega_{opt}$ close to 2, adjusted to maximize the rate of final approach of S to its minimum.

Let us now return to the problem of evaluating the partition function of Eq. (3), and ask whether there is a parametrized, over-relaxation generalization of the Gaussian transition probability of Eq. (10). A simple investigation shows that such a generalization does exist, and is given by

$$w[\{\phi\}_k;\phi(k)\to\phi(k)']=\left[\frac{\omega(2-\omega)}{\pi A_k}\right]^{1/2}\exp\left\{-\left[\frac{A_k}{\omega(2-\omega)}\right][\phi(k)'-\omega C_k-(1-\omega)\phi(k)]^2\right\}, \tag{14a}$$

which can be rewritten as

$$w[\{\phi\}_k;\phi(k)\to\phi(k)']=(\pi A_k\cosh^2\theta)^{-1/2}\exp\{-A_k[\cosh\theta(\phi(k)'-C_k)+\sinh\theta(\phi(k)-C_k)]^2\},\ \omega-1=\tanh\theta. \tag{14b}$$

To verify Eq. (14), we note that it obviously satisfies the normalization condition of Eq. (8a), while since

$$\begin{aligned}A_k[\phi(k)-C_k]^2&+A_k[\cosh\theta(\phi(k)'-C_k)+\sinh\theta(\phi(k)-C_k)]^2\\&=A_k\cosh^2\theta\{[\phi(k)-C_k]^2+[\phi(k)'-C_k]^2\}+2A_k\cosh\theta\sinh\theta[\phi(k)-C_k][\phi(k)'-C_k]\\&=A_k[\phi(k)'-C_k]^2+A_k[\cosh\theta(\phi(k)-C_k)+\sinh\theta(\phi(k)'-C_k)]^2,\end{aligned} \tag{15}$$

it also satisfies the detailed balance condition of Eq. (8b). Hence, the transition probability of Eq. (14) provides an SOR method for the Monte Carlo evaluation of the Euclidean partition function for multiquadratic actions.

To determine whether SOR accelerates the thermalization process, let us analyze in detail the case where the action S is a quadratic (as opposed to a multiquadratic) form, corresponding to an Abelian gauge theory with external sources. Let $\phi(k)^M$ denote the value of the kth-node variable after M complete iterations, let

$$\{\phi\}_k^M=\{\phi(1)^M,\ldots,\phi(k-1)^M,\phi(k+1)^{M-1},\ldots,\phi(N)^{M-1}\}$$

denote the set of node variables which are passive when the kth-node variable is being altered during the Mth-iteration sweep, and let $P[\{\phi\};N(M-1)+k-1]$ be the joint probability distribution of the node variables after $N(M-1)+k-1$ individual node replacements. Then we evidently have

$$P[\{\phi\}_k^M,\phi(k)^M;N(M-1)+k]=\int_{-\infty}^{\infty}d\phi(k)^{M-1}w[\{\phi\}_k^M;\phi(k)^{M-1}\to\phi(k)^M]P[\{\phi\}_k^M,\phi(k)^{M-1};N(M-1)+k-1], \tag{16}$$

which tells us how the joint probability distribution evolves from step to step. Integrating Eq. (16) with respect to $\phi(k)^M$, and using the normalization condition of Eq. (8a), we learn that

$$\int_{-\infty}^{\infty}d\phi(k)^MP[\{\phi\}_k^M,\phi(k)^M;N(M-1)+k]=\int_{-\infty}^{\infty}d\phi(k)^{M-1}P[\{\phi\}_k^M,\phi(k)^{M-1};N(M-1)+k-1], \tag{17}$$

which means that the joint probability distribution for the subset of node variables $\{\phi\}_k$ [with $\phi(k)$ integrated out] is unchanged during the iterative step in which $\phi(k)$ is altered. This in turn implies that the mean value of any node variable $\bar\phi(k)$, defined by

$$\bar\phi(k)=\left[\prod_{i\neq k}\int_{-\infty}^{\infty}d\phi(i)\right]\int_{-\infty}^{\infty}d\phi(k)\phi(k)P[\{\phi\};\ldots], \tag{18}$$

changes only during an iteration step in which $\phi(k)$ is altered, and so is uniquely specified by the notation $\bar\phi(k)^M$, which gives its value after M complete iterations. To study the evolution of the mean values, we multiply Eq. (16) by $\phi(k)^M$ and integrate, giving

$$\begin{aligned}\bar\phi(k)^M&=\left[\prod_{i=1}^{k-1}\int_{-\infty}^{\infty}d\phi(i)^M\right]\int_{-\infty}^{\infty}d\phi(k)^M\phi(k)^M\left[\prod_{i=k+1}^{N}\int_{-\infty}^{\infty}d\phi(i)^{M-1}\right]P[\{\phi\}_k^M,\phi(k)^M;N(M-1)+k]\\&=\left[\prod_{i=1}^{k-1}\int_{-\infty}^{\infty}d\phi(i)^M\right]\left[\prod_{i=k}^{N}\int_{-\infty}^{\infty}d\phi(i)^{M-1}\right]\\&\quad\times\int_{-\infty}^{\infty}d\phi(k)^M\{[\phi(k)^M-\omega C_k-(1-\omega)\phi(k)^{M-1}]_{(1)}+[\omega C_k+(1-\omega)\phi(k)^{M-1}]_{(2)}\}\\&\quad\times w[\{\phi\}_k^M;\phi(k)^{M-1}\to\phi(k)^M]P[\{\phi\}_k^M,\phi(k)^{M-1};N(M-1)+k-1].\end{aligned} \tag{19}$$

Comparing with Eq. (14a), we see that the contribution of the term labeled $[\ ]_{(1)}$ vanishes, while using Eq. (8a) the contribution of the term labeled $[\ ]_{(2)}$ simplifies to give

$$\bar{\phi}(k)^M = \left[\prod_{i=1}^{k-1} \int_{-\infty}^{\infty} d\phi(i)^M\right]\left[\prod_{i=k}^{N} d\phi(i)^{M-1}\right][\omega C_k[\{\phi\}_k^M] + (1-\omega)\phi(k)^{M-1}]\, P[\{\phi\}_k^M, \phi(k)^{M-1}; N(M-1)+k-1]. \qquad (20)$$

Up to this point the analysis is completely general, and applies to multiquadratic as well as quadratic actions. Specializing now to the case of quadratic actions, for which C_k is a linear functional of its arguments, Eq. (20) becomes

$$\bar{\phi}(k)^M = \omega C_k[\{\bar{\phi}\}_k^M] + (1-\omega)\bar{\phi}(k)^{M-1},$$
$$\{\bar{\phi}\}_k^M = \{\bar{\phi}(1)^M, \ldots, \bar{\phi}(k-1)^M, \bar{\phi}(k+1)^{M-1}, \ldots, \bar{\phi}(N)^{M-1}\}. \qquad (21)$$

Thus, under SOR thermalization for a quadratic action, the mean nodal values evolve according to Eq. (21), which is just the difference equation encountered in the SOR minimization of the action S. Since SOR is known to accelerate the minimization process, Eq. (21) implies that it will accelerate convergence of the thermalization process as well. Although the precise statement of Eq. (21) can be made only for quadratic actions, the general conclusion reached here, that SOR accelerates thermalization, is very likely to carry over to the general multiquadratic case as well, much as SOR accelerates the minimization[4] of multiquadratic as well as quadratic actions.

As compared with the conventional[6,7] lattice gauge theory approach, the strategy for evaluating the partition function outlined above may have several advantages. First, since the potentials remain as the variables, there are natural discrete analogs of the gauge potentials and gauge fields, which should permit the study of such questions[13] as the behavior of the effective action for weak fields. Second, since the mean node variables for the Abelian theory thermalize according to the SOR equation encountered in minimizing the discrete action S, and since this equation is just the conventional[9] discrete version of the classical Euclidean field equation derived from the continuum S, the Abelian theory will never give a confining potential for static sources. Thus, if confinement is found in the non-Abelian case, it should not be as an artifact of the discrete procedure. Finally, the SOR method outlined above may well be computationally faster than the lattice gauge theory method, both because of the possibility of acceleration of the thermalization process, and because the Gaussian distribution of Eq. (14a) can be obtained from an array of prestored, normally distributed random numbers by the calculation of a single square root and a relatively small number of arithmetic operations. Detailed numerical experiments will, of course, be needed to see if these conjectured gains are realized in practice.

I wish to thank R. F. Dashen, T. Piran, and S. Samuel for interesting conversations. This work was supported by the U. S. Department of Energy under Grant No. DE-AC02-76ER02220.

[1]For a review, see E. S. Abers and B. W. Lee, Phys. Rep. 9C, 1 (1973).

[2]The Monte Carlo evaluation of fermion functional integrals involves special problems not present in the boson case. For a discussion, see F. Fucito *et al.*, Report No. CERN TH-2960, 1980 (unpublished) and G. De Angelis, D. de Falco, and F. Guerra, Phys. Rev. Rev. D 23, 1747 (1981).

[3]D. Gross, R. Pisarski, and L. Yaffe, Rev. Mod. Phys. 53, 43 (1981).

[4]S. L. Adler and T. Piran, in *High Energy Physics—1980*, proceedings of the XX International Conference, Madison, Wisconsin, edited by L. Durand and L. G. Pondrom (AIP, New York, 1981).

[5]A ϕ^4 term can be formally expressed in multiquadratic form by replacing it with $-\sigma^2 + 2\sigma\phi^2$ with σ an auxiliary variable, but this expression is unbounded from below. (The restriction to multiquadratic actions which are bounded from below, together with the requirement that the action contain no dimensional parameters, automatically selects theories with dynamical scale-invariance breaking. I suspect that this is more than a coincidence.)

[6]K. Wilson, Phys. Rev. D 10, 2445 (1974); for a review, see M. Creutz, Rev. Mod. Phys. 50, 561 (1978). See also J. Kogut and L. Susskind, Phys. Rev. D 11, 395 (1975).

[7]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. D 20, 1915 (1979); M. Creutz, *ibid.* 21, 2308 (1980).

[8]I. M. Singer (unpublished); M. B. Halpern, Phys. Rev. D 19, 517 (1979).

[9]W. F. Ames, *Numerical Methods for Partial Differential Equations* (Academic, New York, 1977), Chap. 3; E. Bauer, O. Betancourt, and P. Garabedian, *A Computational Method in Plasma Physics* (Springer, New York, 1978), Chap. 3.

[10]N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953). For a detailed review, see K. Binder, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 5b.

[11]The detailed-balance condition is the simplest way of satisfying a more general requirement called the homogeneous-state condition. See K. Binder, Ref. 10, for a discussion.

[12]A minimum will exist, in general, if S is bounded from below. If there are several relative minima, one which can be reached by a downhill path from the starting configuration is selected.

[13]S. L. Adler, this issue, Phys. Rev. D 23, 2905 (1981).

PHYSICAL REVIEW D VOLUME 37, NUMBER 2 15 JANUARY 1988

Overrelaxation algorithms for lattice field theories

Stephen L. Adler
*Institute for Scientific Interchange, Villa Gualino-Viale Settimio Severo 65, 10131 Torino, Italy and The Institute for Advanced Study, Princeton, New Jersey 08540**
(Received 1 June 1987)

We study overrelaxation algorithms for the thermalization of lattice field theories with multiquadratic and more general actions. Overrelaxation algorithms are one-parameter generalizations of the heat-bath algorithm which satisfy the detailed-balance condition; the parameter is the relaxation parameter ω, $0<\omega<2$, with $\omega=1$ corresponding to the heat bath. First, we show that the $\omega\to 0$ (extreme underrelaxation) limit of the overrelaxation algorithm is equivalent to the Langevin equation approach. We analyze the thermalization of a free-field action, and show that for $\omega\sim 2$ an overrelaxed Gauss-Seidel algorithm yields a critical slowing down which is independent of wavelength, and has a correlation time which is a factor N smaller than that for an unaccelerated Jacobi iteration, with N the linear dimension of the lattice in lattice units. For a general nonmultiquadratic action, we give a generalized overrelaxation algorithm which satisfies detailed balance with respect to an effective action which is explicitly computable in terms of the original action. In the case of SU(n) lattice gauge theory we use this construction to formulate an overrelaxed algorithm which has exact lattice gauge invariance, and which satisfies detailed balance with respect to an effective action differing from the Wilson action only by terms of relative order a^2 in the continuum limit, with a the lattice spacing.

I. OVERRELAXATION AND ITS RELATION TO THE LANGEVIN APPROACH

The generic lattice field-theory problem is that of evaluating the Euclidean partition function

$$Z=\int d[\phi]e^{-\beta S[\{\phi\}]}\,, \tag{1}$$

with S the Euclidean action on a lattice and with $\int d[\phi]$ an integration over discretized lattice variables. In the Monte Carlo method for evaluating this integral, one generates a Markov chain of configurations $\{\phi_i\}$, $i=1,2,\ldots$ by application of a transition probability $W[\{\phi\}\to\{\phi'\}]$, which is chosen to satisfy the detailed-balance condition

$$e^{-\beta S[\{\phi\}]}W[\{\phi\}\to\{\phi'\}]=e^{-\beta S[\{\phi'\}]}W[\{\phi'\}\to\{\phi\}]\,, \tag{2}$$

as well as normalization and erogodicity conditions.[1] These conditions guarantee that in the limit $i\to\infty$, the ensemble of configurations $\{\phi_i\}$ is distributed according to the equilibrium probability density $\exp(-\beta S[\{\phi\}])$. In what follows, I will refer to the problem of generating such an equilibrium distribution of configurations as the *thermalization problem*. If we now consider the $\beta\to\infty$ (zero-temperature) limit, only the configuration which minimizes S contributes to Eq. (1). Hence the zero-temperature limit of a thermalization algorithm will be an algorithm for the *minimization problem* of finding configurations $\{\bar\phi\}$ which satisfy

$$\frac{\delta}{\delta\phi}S[\{\phi\}]\Big|_{\{\bar\phi\}}=0\,. \tag{3}$$

Conversely, we may expect that by generalizing methods which have been useful in solving the minimization problem, we can get useful algorithms for the thermalization problem.

Following this line of reasoning, a number of years ago I showed[2] that for the special case of multiquadratic actions (which includes[3] the classical Yang-Mills action), the standard Gauss-Seidel overrelaxation algorithm for the minimization problem can be generalized to an overrelaxation algorithm for the thermalization problem. Since this earlier work forms the starting point for the analysis of the present paper, I proceed now to briefly summarize it. A multiquadratic action is one which, for any node variable ϕ_k, can be decomposed as

$$S[\{\phi\}]=S[\{\phi\}_{\neq k},\phi_k]=A_k(\phi_k-C_k)^2+B_k,\quad A_k>0\,, \tag{4}$$

with A_k, B_k, and C_k functions of the remaining node variables

$$\{\phi\}_{\neq k}\equiv\{\phi_i, i\neq k\}\,. \tag{5}$$

A Gauss-Seidel iteration for the minimization problem consists of the successive replacement of each node variable ϕ_k by the value C_k which minimize the action as a function of that single variable, with the other variables $\{\phi\}_{\neq k}$ held fixed. Although this procedure gives the largest single step reduction in S, it is in fact not the most efficient procedure when coherent effects over the entire lattice are taken into account. A better minimization algorithm, which is no more demanding computationally, is the *overrelaxed* Gauss-Seidel algorithm

$$\phi_k\to\phi_k'=\omega C_k+(1-\omega)\phi_k\,, \tag{6}$$

with ω the "relaxation parameter." Convergence is guaranteed provided that S remains nonincreasing at each step, which requires

$$0 \le S[\{\phi\}_{\neq k},\phi_k]-S[\{\phi\}_{\neq k},\phi'_k] = A_k(\phi_k-\phi'_k)^2\left[\frac{2}{\omega}-1\right], \tag{7}$$

giving the restriction

$$0<\omega<2 . \tag{8}$$

When $\omega=1$, Eq. (6) reduces to the Gauss-Seidel prescription $\phi'_k=C_k$, in which the new value ϕ'_k has no memory of the old value ϕ_k. In practice, optimum convergence is obtained by doing several iterations with $\omega=1$, and then doing many iterations with a value of ω close to 2.

Let us now turn to the thermalization problem for the action of Eq. (4). The thermalization analog of the Gauss-Seidel iteration is the heat-bath algorithm, in which a heat bath of temperature β^{-1} is touched in succession to each node variable ϕ_k, with the other variables $\{\phi\}_{\neq k}$ held fixed. In Ref. 2, I showed that the heat-bath algorithm for Eq. (4) admits a one-parameter generalization, analogous to Eq. (6), in which the normalized transition probability W is given by

$$W[\{\phi\}_{\neq k},\phi_k\rightarrow\phi'_k]=\left[\frac{\beta A_k}{\pi\omega(2-\omega)}\right]^{1/2}\exp\left[-\left[\frac{\beta A_k}{\omega(2-\omega)}\right][\phi'_k-\omega C_k-(1-\omega)\phi_k]^2\right]. \tag{9}$$

When $\omega=1$, Eq. (9) reduces to the heat-bath algorithm, since the new values ϕ'_k are distributed according to the equilibrium action and are independent of the old values ϕ_k. To see that Eq. (9) satisfies detailed balance for general ω, let us introduce a hyperbolic angle θ defined by

$$\omega-1=\tanh\theta,\quad \frac{1}{[\omega(2-\omega)]^{1/2}}=\cosh\theta,\quad \frac{1-\omega}{[\omega(2-\omega)]^{1/2}}=-\sinh\theta , \tag{10}$$

in terms of which Eq. (9) takes the form

$$W[\{\phi\}_{\neq k},\phi_k\rightarrow\phi'_k]=(\beta A_k\cosh^2\theta/\pi)^{1/2}\exp\{-\beta A_k[\cosh\theta(\phi'_k-C_k)+\sinh\theta(\phi_k-C_k)]^2\} . \tag{11}$$

Detailed balance now immediately follows from the fact that

$$\begin{aligned}(\phi_k-C_k)^2+[\cosh\theta(\phi'_k-C_k)+\sinh\theta(\phi_k-C_k)]^2&=\cosh^2\theta[(\phi_k-C_k)^2+(\phi'_k-C_k)^2]\\&\quad+2\cosh\theta\sinh\theta(\phi_k-C_k)(\phi'_k-C_k)\\&=\text{symmetric in }\phi_k,\phi'_k .\end{aligned} \tag{12}$$

Since the transition probability of Eq. (9) is a Gaussian, it can be conveniently represented as a stochastic difference equation. Let n by a fictitious "time" index which increases by one for each update of the entire lattice, and let $\eta_{n,k}$ be a set of Gaussian noise variables distributed according to

$$W[\{\eta\}]=\prod_{n,k}\left[\frac{1}{\sqrt{4\pi}}e^{-\eta_{n,k}{}^2/4}\right], \tag{13}$$

and hence which obey

$$\langle\eta_{n,k}\eta_{n',k'}\rangle_\eta=2\delta_{n,n'}\delta_{k,k'} . \tag{14}$$

Writing $\phi_k\equiv\phi_k^n$, $\phi'_k\equiv\phi_k^{n+1}$, Eq. (9) is evidently equivalent to

$$\phi_k^{n+1}-\phi_k^n=-\omega(\phi_k^n-C_k)-\left[\frac{\omega(2-\omega)}{4\beta A_k}\right]^{1/2}\eta_{n,k} , \tag{15}$$

with C_k and A_k functions of the ϕ_i^{n+1} for those nodes which precede ϕ_k, and of ϕ_i^n for those nodes which follow ϕ_k, in the sweep of the lattice. [This just corresponds to the fact that an updating of the whole lattice is accomplished by the *successive* application of the transition probability of Eq. (9) to each node of the lattice, in some specified sweep order.] Let us now rewrite Eq. (15) by using the fact that ϕ_k-C_k is proportional to $\partial S/\partial\phi_k$,

$$2\beta A_k(\phi_k^n-C_k)=\beta\frac{\partial S}{\partial\phi_k}[\{\phi_{i<k}^{n+1}\},\{\phi_{i\ge k}^n\}]\equiv\beta\frac{\partial S}{\partial\phi_k} , \tag{16}$$

and by defining ϵ_k according to

$$\frac{\omega}{2\beta A_k}=\epsilon_k[\{\phi_{i<k}^{n+1}\},\{\phi_{i\ge k}^n\}]\equiv\epsilon_k , \tag{17}$$

giving

$$\phi_k^{n+1}-\phi_k^n=-\epsilon_k\beta\frac{\partial S}{\partial\phi_k}-\left[1-\frac{\omega}{2}\right]^{1/2}\epsilon_k{}^{1/2}\eta_{n,k} . \tag{18}$$

Apart from the extra factor of $(1-\omega/2)^{1/2}$, which approaches unity as $\omega\rightarrow0$, Eq. (18) is just the discrete form of the Langevin equation with variable step size ϵ_k and a Gauss-Seidel interpretation of $\partial S/\partial\phi$, and approaches the corresponding Langevin stochastic differential equa-

tion as $\epsilon_k \propto \omega \rightarrow 0$. Hence the Langevin equation approach corresponds to the *extreme underrelaxation* limit[4] of the overrelaxation algorithm of Eq. (9).

II. CRITICAL SLOWING DOWN FOR A FREE-FIELD ACTION

In this section we give a detailed theoretical analysis of the performance of the overrelaxed minimization and thermalization algorithms, motivated by the fact that numerical studies by Whitmer,[5] Creutz,[4] and Brown and Woch[6] suggest that overrelaxation can improve the correlation time, as well as the speed of thermalization, in Monte Carlo simulations. We consider for simplicity the case of a single massless scalar free field ϕ in d dimensions; the inclusion of interaction and mass terms is not expected[7] to change the qualitative conclusions reached below. The node variable is thus

$$\phi_{i_1,\ldots,i_d} \tag{19}$$

and the action is taken as

$$S=\sum_{i_1,\ldots,i_d=1}^{N} \tfrac{1}{2}[(\phi_{i_1+1,i_2,\ldots,i_d}-\phi_{i_1,i_2,\ldots,i_d})^2 +(\phi_{i_1,i_2+1,\ldots,i_d}-\phi_{i_1,i_2,\ldots,i_d})^2 + \cdots +(\phi_{i_1,\ldots,i_d+1}-\phi_{i_1,\ldots,i_d})^2], \tag{20}$$

with homogeneous (Dirichlet or Neumann) boundary conditions applied at the edges of the lattice.[8] Introducing the notation

$$\phi(i_\mu \pm 1) \equiv \phi_{i_1,\ldots,i_d} \,|_{i_{l\neq\mu}\text{ fixed},\ i_\mu \rightarrow i_\mu \pm 1} , \tag{21}$$

we can now write the dependence of S on a given node $\phi_I \equiv \phi_{i_1,\ldots,i_d}$ as

$$S=\tfrac{1}{2}\sum_{\mu=1}^{d} \{[\phi(i_\mu+1)-\phi_I]^2+[\phi(i_\mu-1)-\phi_I]^2\}+\bar{S} , \quad \bar{S} \text{ independent of } \phi_I . \tag{22}$$

From Eqs. (4) and (11) of Sec. I, we see that an overrelaxed transition probability for the update $\phi_I \rightarrow \phi_I'$ can be constructed as

$$W[\phi_I \rightarrow \phi_I'] = \mathcal{N} \exp\left[-\tfrac{1}{2}\beta \sum_{\mu=1}^{d} (\{\cosh\theta[\phi(i_\mu+1)-\phi_I']+\sinh\theta[\phi(i_\mu+1)-\phi_I]\}^2 + \{\cosh\theta[\phi(i_\mu-1)-\phi_I']+\sinh\theta[\phi(i_\mu-1)-\phi_I]\}^2) \right] , \tag{23}$$

with the normalization constant $\mathcal{N}$ independent of ϕ_I and ϕ_I'. Rewriting the ϕ_I' dependence by completing the square, and then substituting Eq. (10), Eq. (23) can be reexpressed as

$$\begin{aligned} W[\phi_I \rightarrow \phi_I'] &= \mathcal{N} \exp\left\{ -\frac{\beta}{4d} \left[\cosh\theta \left[2d\phi_I' - \sum_{\mu=1}^{d} [\phi(i_\mu+1)+\phi(i_\mu-1)] \right] + \sinh\theta \left[2d\phi_I - \sum_{\mu=1}^{d} [\phi(i_\mu+1)+\phi(i_\mu-1)] \right] \right]^2 + \phi_I,\phi_I'\text{-independent} \right\} \\ &= \mathcal{N} \exp\left[-\frac{1}{4} \left[\frac{4\beta}{d\omega(2-\omega)} \right] \left[d\phi_I' - (1-\omega)d\phi_I - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [\phi(i_\mu+1)+\phi(i_\mu-1)] \right]^2 + \phi_I,\phi_I'\text{-independent} \right] . \end{aligned} \tag{24}$$

Applying the procedure of Eqs. (13)–(18), Eq. (24) can be rewritten as a Gauss-Seidel stochastic difference equation

$$\begin{aligned} & d\phi_I^{n+1} - (1-\omega)d\phi_I^n - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [\phi^n(i_\mu+1)+\phi^{n+1}(i_\mu-1)] = -\sigma\eta_{I,n} , \\ & \langle \eta_{I,n}\eta_{I',n'} \rangle_\eta = 2\delta_{I,I'}\delta_{n,n'}, \quad \sigma = \left[\frac{d\omega(2-\omega)}{4\beta} \right]^{1/2} , \quad \delta_{I,I'} \equiv \delta_{i_1,i_1'} \cdots \delta_{i_d,i_d'} . \end{aligned} \tag{25}$$

It will be informative, in what follows, to also analyze the corresponding Jacobi stochastic difference equation, in which the old values $\phi^n(i_\mu-1)$ are used for the earlier nodes in the sweep, instead of the updated values $\phi^{n+1}(i_\mu-1)$,

$$d\phi_I^{n+1} - (1-\omega)d\phi_I^n - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [\phi^n(i_\mu+1)+\phi^{n+1}(i_\mu-1)] = -\sigma\eta_{I,n} , \tag{26}$$

Although Eq. (26) is not equivalent to the iteration of an algorithm which satisfies detailed balance with the action of Eq. (20), it has been extensively studied by Batrouni *et al.*,[9] and so furnishes a useful point of comparison.

To solve Eqs. (25) and (26), we proceed by introducing a Green's function

$$G^{n,n'}_{i_1,\ldots,i_d;i'_1,\ldots,i'_d} \equiv G^{n,n'}_{I,I'} , \tag{27}$$

which satisfies the stochastic difference equation,

Gauss-Seidel case:

$$dG^{n+1,n'}_{I,I'} - (1-\omega)dG^{n,n'}_{I,I'} - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [G^{n,n'}_{I'}(i_\mu+1) + G^{n+1,n'}_{I'}(i_\mu-1)] = \delta_{I,I'}\delta_{n,n'} , \tag{28a}$$

Jacobi case:

$$dG^{n+1,n'}_{I,I'} - (1-\omega)dG^{n,n'}_{I,I'} - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [G^{n,n'}_{I'}(i_\mu+1) + G^{n,n'}_{I'}(i_\mu-1)] = \delta_{I,I'}\delta_{n,n'} \tag{28b}$$

and the boundary condition

$$G^{0,n'}_{I,I'} = 0 . \tag{29}$$

Then the solution of Eqs. (25) and (26) can be written as

$$\phi^n_I = -\sum_{I',n'} G^{n,n'}_{I,I'} \sigma \eta_{I',n'} + \bar{\phi}^n_I , \tag{30}$$

where $\bar{\phi}^n_I$ is the solution of the $\sigma=0$ (noise-free, or zero temperature) iteration,

Gauss-Seidel case:

$$d\bar{\phi}^{n+1}_I - (1-\omega)d\bar{\phi}^n_I - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [\bar{\phi}^n(i_\mu+1) + \bar{\phi}^{n+1}(i_\mu-1)] = 0 , \tag{31a}$$

Jacobi case:

$$d\bar{\phi}^{n+1}_I - (1-\omega)d\bar{\phi}^n_I - \tfrac{1}{2}\omega \sum_{\mu=1}^{d} [\bar{\phi}^n(i_\mu+1) + \bar{\phi}^n(i_\mu-1)] = 0 \tag{31b}$$

with the initial condition $\bar{\phi}^0_I = \phi^0_I$. Since Eq. (30) implies that

$$\langle \phi^n_I \rangle_\eta = \bar{\phi}^n_I , \tag{32}$$

we see that introduction of the Green's function has permitted us to separate ϕ^n_I into a mean value term and individual noise contributions. The rapidity of thermalization is determined by the rate of decay of $\bar{\phi}^n_I$ with n, while the correlation time (the number of updates required to evolve from one thermalized configuration to an independent one) is determined by the rate of decay of $G^{n,n'}_{I,I'}$ with n. For a general Monte Carlo calculation the thermalization and the correlation times are different, but they will turn out to be equal for the overrelaxed quadratic action case studied in this section.

Since we are really interested only in the asymptotic limit of small mesh spacings or large lattices, we do not attempt to solve the difference equations (28) and (31) directly. (An alternative method, working directly from the iteration matrix for the difference equations, and yielding similar conclusions, has been given by Goodman and Sokal.[10]) Instead we follow the method of Garabedian[11] and convert the discrete equations to an equivalent continuum problem, for which the corresponding partial differential equations can be solved by standard methods. Let us denote the mesh spacing by a and introduce continuum variables x_μ, t by the correspondence

$$\begin{gathered} x_\mu \leftrightarrow a i_\mu, \quad \int dx_\mu \leftrightarrow a \sum_{i_\mu}, \quad d/dx_\mu \leftrightarrow a^{-1}\Delta_{i_\mu} , \\ t \leftrightarrow an, \quad \int dt \leftrightarrow a \sum_n, \quad d/dt \leftrightarrow a^{-1}\Delta_t , \\ a^{d+1}\delta(x_1-x'_1)\cdots\delta(x_d-x'_d)\delta(t-t') \leftrightarrow \delta_{I,I'}\delta_{n,n'} , \end{gathered} \tag{33}$$

with Δ the finite difference operator. Treating first the Jacobi iteration case, we rewrite Eqs. (28b) and (31b) as

$$\begin{gathered} a^{-1}da^{-1}(G^{n+1,n'}_{I,I'} - G^{n,n'}_{I,I'}) - \tfrac{1}{2}\omega a^{-2} \sum_{\mu=1}^{d} [G^{n,n'}_{I'}(i_\mu+1) + G^{n,n'}_{I'}(i_\mu-1) - 2G^{n,n'}_{I,I'}] = a^{d-1}a^{-d-1}\delta_{I,I'}\delta_{n,n'} , \\ a^{-1}da^{-1}(\bar{\phi}^{n+1}_I - \bar{\phi}^n_I) - \tfrac{1}{2}\omega a^{-2} \sum_{\mu=1}^{d} [\bar{\phi}^n(i_\mu+1) + \bar{\phi}^n(i_\mu-1) - 2\bar{\phi}^n_I] = 0 . \end{gathered} \tag{34}$$

Making the correspondence

$$G^{n,n'}_{I,I'} \leftrightarrow G(x,x';t,t'), \quad \bar{\phi}^n_I \leftrightarrow \bar{\phi}(x,t) \tag{35}$$

and referring to Eq. (33), we see that Eqs. (34) are the discrete analogs of the continuum parabolic partial differential equations

$$\begin{gathered} \frac{2d}{\omega a}\frac{\partial}{\partial t}G(x,x';t,t') - \sum_{\mu=1}^{d} \frac{\partial^2}{\partial x_\mu{}^2}G(x,x';t,t') = \frac{2}{\omega}a^{d-1}\delta(x_1-x'_1)\cdots\delta(x_d-x'_d)\delta(t-t') , \\ \frac{2d}{\omega a}\frac{\partial}{\partial t}\bar{\phi}(x,t) - \sum_{\mu=1}^{d} \frac{\partial^2}{\partial x_\mu{}^2}\bar{\phi}(x,t) = 0 , \end{gathered} \tag{36}$$

with the initial conditions

$$G(x,x';0,t')=0,\quad t'>0,\quad \bar{\phi}(x,0)=\text{smooth interpolation of } \phi_I^0 \ . \tag{37}$$

Turning next to the Gauss-Seidel iteration case, we follow the Garabedian analysis and anticipate the fact that the optimum ω is related to the mesh spacing a by

$$\omega=\frac{2}{1+Ca}\ , \tag{38}$$

with C a constant of order unity. Substituting Eq. (38) into Eqs. (28a) and (31a), these can be rewritten in the form

$$\begin{aligned} &dCa^{-1}(G_{I,I'}^{n+1,n'}-G_{I,I'}^{n,n'})-a^{-2}\sum_{\mu=1}^{d}[G_{I'}^{n,n'}(i_\mu+1)+G_{I'}^{n,n'}(i_\mu-1)-2G_{I,I'}^{n,n'}]\\ &\qquad+a^{-2}\sum_{\mu=1}^{d}\{G_{I'}^{n+1,n'}(i_\mu)-G_{I'}^{n+1,n'}(i_\mu-1)-[G_{I'}^{n,n'}(i_\mu)-G_{I'}^{n,n'}(i_\mu-1)]\}=\frac{2}{\omega}a^{d-1}a^{-d-1}\delta_{I,I'}\delta_{n,n'}\ ,\\ &dCa^{-1}(\bar{\phi}_I^{n+1}-\bar{\phi}_I^{n})-a^{-2}\sum_{\mu=1}^{d}[\bar{\phi}^{n}(i_\mu+1)+\bar{\phi}^{n}(i_\mu-1)-2\bar{\phi}_I^{n}]+a^{-2}\sum_{\mu=1}^{d}\{\bar{\phi}^{n+1}(i_\mu)-\bar{\phi}^{n+1}(i_\mu-1)\\ &\qquad-[\bar{\phi}^{n}(i_\mu)-\bar{\phi}^{n}(i_\mu-1)]\}=0\ . \end{aligned} \tag{39}$$

Again making the correspondence of Eq. (35), we see that Eqs. (39) are the discrete analogs of the continuum hyperbolic partial differential equations:

$$\begin{aligned} &dC\frac{\partial}{\partial t}G(x,x';t,t')-\sum_{\mu=1}^{d}\frac{\partial^2}{\partial x_\mu{}^2}G(x,x';t,t')+\sum_{\mu=1}^{d}\frac{\partial^2}{\partial t\partial x_\mu}G(x,x';t,t')=\frac{2}{\omega}a^{d-1}\delta(x_1-x_1')\cdots\delta(x_d-x_d')\delta(t-t')\ ,\\ &dC\frac{\partial}{\partial t}\bar{\phi}(x,t)-\sum_{\mu=1}^{d}\frac{\partial^2}{\partial x_\mu{}^2}\bar{\phi}(x,t)+\sum_{\mu=1}^{d}\frac{\partial^2}{\partial t\partial x_\mu}\bar{\phi}(x,t)=0\ , \end{aligned} \tag{40}$$

with initial conditions[12] as in Eq. (37). Now making the change of variable

$$s=t+\tfrac{1}{2}\sum_{\mu=1}^{d}x_\mu\ , \tag{41}$$

some straightforward algebra shows that Eq. (40) is transformed into the canonical hyperbolic form

$$\begin{aligned} &dC\frac{\partial}{\partial s}G(x,x';s,s')+\frac{d}{4}\frac{\partial^2}{\partial s^2}G(x,x';s,s')-\sum_{\mu=1}^{d}\frac{\partial^2}{\partial x_\mu{}^2}G(x,x';s,s')=\frac{2}{\omega}a^{d-1}\delta(x_1-x_1')\cdots\delta(x_d-x_d')\delta(s-s')\ ,\\ &dC\frac{\partial}{\partial s}\bar{\phi}(x,s)+\frac{d}{4}\frac{\partial^2}{\partial s^2}\bar{\phi}(x,s)-\sum_{\mu=1}^{d}\frac{\partial^2}{\partial x_\mu{}^2}\bar{\phi}(x,s)=0\ , \end{aligned} \tag{42}$$

with the boundary conditions

$$\begin{aligned} &G(x,x';s,s')=0\ ,\\ &\bar{\phi}(x,s)=\text{smooth interpolation of } \phi_I^0 \end{aligned} \tag{43a}$$

imposed on the surface

$$s=\tfrac{1}{2}\sum_{\mu=1}^{d}x_\mu\ . \tag{43b}$$

Let us proceed now to solve Eqs. (36) and (42) by separation of variables. Let $\psi_m(x)$ be a complete set of eigenfunctions of the d-dimensional Laplace operator

$$\nabla^2=\sum_{\mu=1}^{d}\frac{\partial^2}{\partial x_\mu{}^2}\ , \tag{44}$$

subject to homogeneous boundary conditions on the edge of the cube $0\le x_\mu \le L$ which bounds the lattice.

We assume that the boundary conditions are such that there are no normalizable zero modes. (This implies that the corresponding Laplace equation with inhomogeneous boundary conditions has a unique solution.) Then we have

$$\begin{aligned} &\nabla^2\psi_m=-k_m{}^2\psi_m\ ,\\ &\delta(x_1-x_1')\cdots\delta(x_d-x_d')=\sum_m \psi_m(x)\psi_m^*(x')\ , \end{aligned} \tag{45}$$

with the minimum eigenvalue k_1 of order L^{-1}. In the Jacobi iteration case, we expand

$$\begin{aligned} &G(x,x';t,t')=\sum_m G_m(x';t,t')\psi_m(x)\ ,\\ &\bar{\phi}(x,t)=\sum_m \bar{\phi}_m(t)\psi_m(x)\ , \end{aligned} \tag{46}$$

with the coefficients G_m and $\bar{\phi}_m$ obeying the differential equations

$$\frac{2d}{\omega a}\frac{d}{dt}G_m(x';t,t')+k_m{}^2G_m(x';t,t') = \frac{2}{\omega}a^{d-1}\psi_m^*(x')\delta(t-t') ,$$
$$\frac{2d}{\omega a}\frac{d}{dt}\bar{\phi}_m(t)+k_m{}^2\bar{\phi}_m(t)=0 . \tag{47}$$

Expanding the initial condition on $\bar{\phi}$ as

$$\bar{\phi}(x,0)=\sum_m \phi_m^{(0)}\psi_m(x) , \tag{48}$$

Eqs. (47) are readily integrated to give

$$G(x,x';t,t')=\frac{a^d}{d}\sum_m \psi_m(x)\psi_m^*(x')e^{-\lambda_m(t-t')}\theta(t-t') ,$$
$$\bar{\phi}(x,t)=\sum_m \phi_m^{(0)}\psi_m(x)e^{-\lambda_m t}, \quad \lambda_m=\frac{\omega a k_m{}^2}{2d} . \tag{49}$$

We turn next to the Gauss-Seidel iteration case. We now expand

$$G(x,x';s,s')=\sum_m G_m(x';s,s')\psi_m(x) ,$$
$$\bar{\phi}(x,s)=\sum_m \bar{\phi}_m(s)\psi_m(x) , \tag{50}$$

with the coefficients G_m and $\bar{\phi}_m$ obeying the differential equations

$$dC\frac{d}{ds}G_m(x';s,s')+\frac{d}{4}\frac{d^2}{ds^2}G_m(x';s,s') +k_m{}^2G_m(x';s,s')=\frac{2}{\omega}a^{d-1}\psi_m^*(x')\delta(s-s') ,$$
$$dC\frac{d}{ds}\bar{\phi}_m(s)+\frac{d}{4}\frac{d^2}{ds^2}\bar{\phi}_m(s)+k_m{}^2\bar{\phi}_m(s)=0 . \tag{51}$$

The general solution for $\bar{\phi}_m(s)$ has the form first given by Garabedian:[11]

$$\bar{\phi}_m(s)=a_m e^{-p_m s}+b_m e^{-q_m s} ,$$
$$p_m=2[C-(C^2-k_m{}^2/d)^{1/2}] , \tag{52}$$
$$q_m=2[C+(C^2-k_m{}^2/d)^{1/2}] ,$$

with the coefficients a_m, b_m implicitly (but not explicitly) determined by matching to the initial condition on $\bar{\phi}$ of Eq. (43). The values of a_m, b_m in fact do not matter; all we need for what follows is that $\bar{\phi}$ decays as a function of time at least as fast as

$$\exp[-t \min_m(\mathrm{Re}p_m, \mathrm{Re}q_m)] . \tag{53}$$

To solve the equation for G_m, we write

$$G_m(x';s,s')=\begin{cases} G_m^>(x';s,s'), & s>s' ,\\ G_m^<(x';s,s'), & s<s' ,\end{cases} \tag{54}$$

with

$$G_m^{>,<}(x';s,s')=a_m^{>,<}(x';s')e^{-p_m s}+b_m^{>,<}(x';s')e^{-q_m s} . \tag{55}$$

Continuity across $s=s'$ requires

$$G_m^>(x';s',s')=G_m^<(x';s',s') , \tag{56}$$

while the δ function on the right-hand side of Eq. (51) requires the first derivative discontinuity to be

$$\frac{d}{4}\frac{d}{ds}[G_m^>(x';s,s')-G_m^<(x';s,s')]\bigg|_{s=s'}=\frac{2}{\omega}a^{d-1}\psi_m^*(x') . \tag{57}$$

To solve these, let us make the ansatz

$$G_m^< = 0 \tag{58}$$

and then show that this does in fact satisfy the boundary condition of Eq. (43). Assuming Eq. (58), a little algebra shows that Eqs. (56) and (57) are satisfied by

$$G_m^>(x';s,s')=\frac{4}{d}\frac{2}{\omega}a^{d-1}\psi_m^*(x')\frac{e^{p_m(s'-s)}-e^{q_m(s'-s)}}{q_m-p_m} , \tag{59}$$

as can be verified by inspection. Hence $G(x,x';s,s')$ is given by

$$G(x,x';s,s')=\frac{4}{d}\frac{2}{\omega}a^{d-1} \times\sum_m \psi_m(x)\psi_m^*(x') \times\frac{e^{p_m(s'-s)}-e^{q_m(s'-s)}}{q_m-p_m}\theta(s-s') \tag{60}$$

and corresponds to taking a solution to the hyperbolic equation Eq. (42) which has support only inside the forward light cone:

$$\left[\sum_{\mu=1}^{d}(x_\mu-x'_\mu)^2\right]^{1/2}\le\frac{2}{d^{1/2}}(s-s') . \tag{61}$$

Now the Schwartz inequality implies

$$\pm\sum_{\mu=1}^{d}(x_\mu-x'_\mu)\le\left[\sum_{\mu=1}^{d}1\right]^{1/2}\left[\sum_{\mu=1}^{d}(x_\mu-x'_\mu)^2\right]^{1/2} =d^{1/2}\left[\sum_{\mu=1}^{d}(x_\mu-x'_\mu)^2\right]^{1/2} \tag{62}$$

and combining the inequalities of Eqs. (61) and (62), we see that inside the forward light cone we have the inequalities[13]

$$0\le s-s'\mp\tfrac{1}{2}\sum_{\mu=1}^{d}(x_\mu-x'_\mu)=\begin{cases} t-t' \\ t-t'+\sum_{\mu=1}^{d}(x_\mu-x'_\mu) .\end{cases} \tag{63}$$

Thus for $t'>0$, the surface $t=0$ lies entirely *outside* the forward light cone, and hence the initial condition that G vanish at $t=0$ is satisfied by Eq. (60), even though the differential equation is not separable in the x,t coordinate system. From Eq. (60) we learn that G also decays

as a function of time at least as fast as Eq. (53).

Let us now determine (following again Ref. 7) the value of C which maximizes the decay exponent

$$\lambda_{GS}=\min_m(\mathrm{Re}p_m,\mathrm{Re}q_m)\,, \tag{64}$$

where GS denotes Gauss-Seidel. For $k_m{}^2$ large enough so that $k_m{}^2/d\ge C^2$, we have

$$\mathrm{Re}p_m=\mathrm{Re}q_m=2C\,. \tag{65}$$

On the other hand, for values of $k_m{}^2$ small enough so that $(C^2-k_m{}^2/d)^{1/2}$ is real, we have

$$\min(\mathrm{Re}p_m,\mathrm{Re}q_m)=p_m\ge p_1=2[C-(C^2-k_1{}^2/d)^{1/2}]\,. \tag{66}$$

Hence

$$\min_m(\mathrm{Re}p_m,\mathrm{Re}q_m) =\min\{2C,2[C-\mathrm{Re}(C^2-k_1{}^2/d)^{1/2}]\}\,, \tag{67}$$

and this expression is maximized for

$$C_{\mathrm{opt}}=k_1/d^{1/2}\,. \tag{68}$$

At the optimum C we have

$$\lambda_{GS}=\mathrm{Re}p_m=\mathrm{Re}q_m=2C_{\mathrm{opt}}=2k_1/d^{1/2} \tag{69}$$

and all modes have the *same* time decay exponent. By contrast, for the Jacobi iteration the decay exponent [Eq. (49)] varies quadratically with wave number

$$\lambda_m=\omega_J a k_m{}^2/d \tag{70}$$

and becomes very small at the largest wavelengths, giving for the most slowly decaying mode

$$\lambda_J=\omega_J a k_1{}^2/d\,. \tag{71}$$

Comparing Eqs. (69) and (71) we have

$$\frac{\lambda_{GS}}{\lambda_J}=\frac{2}{\omega_J}d^{1/2}\frac{1}{k_1 a}\,. \tag{72a}$$

Since $k_1\sim L^{-1}$ and $L/a=N$, with N the dimension of the lattice in lattice units, we get our fundamental result

$$\frac{\lambda_{GS}}{\lambda_J}\sim\frac{2}{\omega_J}d^{1/2}N\,. \tag{72b}$$

Hence the overrelaxed Gauss-Seidel algorithm dramatically improves both the rapidity of thermalization and the correlation time as compared with the Jacobi algorithm, and makes critical slowing down independent of wave length. This improvement becomes even more pronounced when compared with Langevin-Jacobi procedures, for which (as shown in Sec. I) one has $\omega_J\ll 1$.

We conclude this section with two checks on the analysis given above. First, the Jacobi case analyzed above is just a continuum version of the model for the correlation length studied by Batrouni *et al.*[9] In units with $a=1$, they find

$$N_c\sim\frac{1}{\bar\epsilon(p^2+m^2)}\,, \tag{73}$$

which with the correspondences $\bar\epsilon\sim\omega_J$ (cf. Sec. I) and $p^2+m^2\sim k_m{}^2$ becomes

$$N_c\sim\frac{1}{\omega_J k_m^2}\sim\lambda_m{}^{-1}\,, \tag{74}$$

and so our result agrees with theirs. Second, as a check on the reasoning leading to Eq. (60), we have explicitly evaluated the time dependence of the Gauss-Seidel Green's function for the $L\to\infty$ limit in which the ψ_m are infinite-space mode functions. Details of this calculation are given in the Appendix; the result is

$$\begin{aligned}G(x,x';t,t')&=\frac{4}{d}\frac{2}{\omega}\frac{a^{d-1}}{(2\pi)^d}\int d^d l\, e^{il\cdot(x-x')}g(l,t-t')\,,\\ g(l^{\parallel},l^{\perp},t)&=\frac{1}{4(C+il^{\parallel}/d^{1/2})}\\ &\quad\times\exp\left[-t\frac{(l^{\parallel})^2+(l^{\perp})^2}{(l^{\parallel})^2+C^2d}\times(C-il^{\parallel}/d^{1/2})\right]\theta(t)\,,\end{aligned} \tag{75}$$

with $l^{\parallel}$ and $l^{\perp}$ the components of l parallel and perpendicular to the fixed vector $(1,1,\ldots,1)$. The presence of the factor $\theta(t-t')$ implies that $G(x,x';t,t')$ vanishes at $t=0$ for $t'>0$, and so Eq. (60) does indeed satisfy the initial condition of Eq. (37). For $C=k_1/d^{1/2}$, Eq. (75) implies that the decay exponent is $\sim k_1$ for wave numbers $|l|$ larger than k_1, in agreement with Eq. (69). [For wave numbers $|l|$ smaller than k_1 Eq. (75) is no longer relevant, since the difference between infinite space and finite box mode functions becomes significant.]

III. A GENERALIZED OVERRELAXATION ALGORITHM, AND APPLICATION TO SU(n) LATTICE FIELD AND GAUGE THEORY

The results of Sec. II indicate that overrelaxation should be of computational value for the thermalization problem, and so we proceed next to construct overrelaxation algorithms for the Yang-Mills action (which, as noted above, is multiquadratic in the components of the gauge potential), and for the Wilson lattice gauge action (which is not multiquadratic). The construction employs the following generalization of the overrelaxation algorithm of Sec. I: Consider a field theory with field variables which can be divided into two disjoint classes $\{\phi\},\{\psi\}$, with functional integration measure

$$d\mu=\prod_1^{N_\phi}\int d\phi\prod_1^{N_\psi}\int d\psi \tag{76a}$$

and with the general (nonmultiquadratic) action

$$S=S_1[\{\phi\},\{\psi\}]+S_2[\{\psi\}]\,. \tag{76b}$$

For this theory, consider an updating $\{\phi\}\to\{\phi'\}$ in which only the $\{\phi\}$ variables (or some subset of them) is changed, and let $\bar S[\{\phi\},\{\phi'\},\{\psi\};\theta]$ be any auxiliary functional of the indicated field variables and the relaxation parameter θ which is symmetrical under the interchange $\{\phi\}\leftrightarrow\{\phi'\}$. For this updating, we take the transition probability to be

$$W[\{\phi\}\to\{\phi'\}]=\mathcal{N}[\{\phi\},\{\psi\};\theta]\exp(-\beta\cosh^2\theta S_1[\{\phi'\},\{\psi\}]$$
$$-\beta\sinh^2\theta S_1[\{\phi\},\{\psi\}]-\beta\tilde{S}[\{\phi\},\{\phi'\},\{\psi\};\theta]) \tag{77a}$$

with the normalization $\mathcal{N}[\{\phi\},\{\psi\};\theta]$ given by

$$\mathcal{N}^{-1}[\{\phi\},\{\psi\};\theta]=\prod_1^{N_\phi}\int d\phi'\exp(-\beta\cosh^2\theta S_1[\{\phi'\},\{\psi\}]-\beta\sinh^2\theta S_1[\{\phi\},\{\psi\}]-\beta\tilde{S}[\{\phi\},\{\phi'\},\{\psi\};\theta]) . \tag{77b}$$

Then W satisfies detailed balance with respect to the effective action

$$S_{\rm eff}[\{\phi\},\{\psi\};\theta]=S[\{\phi\},\{\psi\}]+\beta^{-1}\ln(\mathcal{N}[\{\phi\},\{\psi\};\theta]/\bar{\mathcal{N}}[\{\psi\};\theta]) ,$$
$$\bar{\mathcal{N}}[\{\psi\};\theta]=\prod_1^{N_\phi}\int d\phi\,\mathcal{N}[\{\phi\},\{\psi\};\theta]\Big/\prod_1^{N}\int d\phi . \tag{78}$$

The proof follows directly from the fact that

$$W[\{\phi\}\to\{\phi'\}]\exp(-\beta S[\{\phi\},\{\psi\}]-\ln(\mathcal{N}[\{\phi\},\{\psi\};\theta]/\bar{\mathcal{N}}[\{\psi\};\theta]))$$
$$=\bar{\mathcal{N}}[\{\psi\};\theta]\exp(-\beta\cosh^2\theta(S_1[\{\phi'\},\{\psi\}]+S_1[\{\phi\},\{\psi\}])-\beta\tilde{S}[\{\phi\},\{\phi'\},\{\psi\};\theta]-\beta S_2[\{\psi\}])$$
$$=\text{symmetrical in } \{\phi\},\{\phi'\} . \tag{79}$$

The multiquadratic case of the generalized algorithm is recovered by taking

$$S_1[\{\phi\},\{\psi\}]=\sum_{ij}L_i[\{\phi\},\{\psi\}]A_{ij}[\{\psi\}]L_j[\{\phi\},\{\psi\}] \tag{80}$$

with the L_i linear functionals of the subset of variables $\{\phi\}$. The overrelaxation algorithm of Secs. I and II then corresponds to the choice of auxiliary functional

$$\tilde{S}[\{\phi\},\{\phi'\},\{\psi\};\theta]=\sinh\theta\cosh\theta\sum_{ij}(L_i[\{\phi'\},\{\psi\}]A_{ij}[\{\psi\}]L_j[\{\phi\},\{\psi\}]$$
$$+L_i[\{\phi\},\{\psi\}]A_{ij}[\{\psi\}]L_j[\{\phi'\},\{\psi\}]) , \tag{81}$$

for which the transition probability W of Eq. (77a) becomes

$$W[\{\phi\}\to\{\phi'\}]=\mathcal{N}\exp\left[-\beta\sum_{ij}(\cosh\theta L_i[\{\phi'\},\{\psi\}]+\sinh\theta L_i[\{\phi\},\{\psi\}])\right.$$
$$\left.\times A_{ij}[\{\psi\}](\cosh\theta L_j[\{\phi'\},\{\psi\}]+\sinh\theta L_j[\{\phi\},\{\psi\}])\right] ; \tag{82a}$$

in terms of a relaxation parameter ω related to θ as in Eq. (10), this can also be written as

$$W[\{\phi\}\to\{\phi'\}]=\mathcal{N}\exp\left[-\frac{\beta}{\omega(2-\omega)}\sum_{ij}(L_i[\{\phi'\},\{\psi\}]-(1-\omega)L_i[\{\phi\},\{\psi\}])\right.$$
$$\left.\times A_{ij}[\{\psi\}](L_j[\{\phi'\},\{\psi\}]-(1-\omega)L_j[\{\phi\},\{\psi\}])\right] . \tag{82b}$$

Since by the linearity of L we have

$$\cosh\theta L_i[\{\phi'\},\{\psi\}]+\sinh\theta L_i[\{\phi\},\{\psi\}]=L_i[\{\cosh\theta\phi'+\sinh\theta\phi\},\{\psi\}] , \tag{83}$$

the normalization is now given by

$$\mathcal{N}^{-1}=\prod_1^{N_\phi}\int d\phi'\exp(-\beta L_i[\{\cosh\theta\phi'+\sinh\theta\phi\},\{\psi\}]A_{ij}[\{\psi\}]L_j[\{\cosh\theta\phi'+\sinh\theta\phi\},\{\psi\}])$$
$$=(\cosh\theta)^{-N_\phi}\prod_1^{N_\phi}\int d\phi'\exp(-\beta L_i[\{\phi'\},\{\psi\}]A_{ij}[\{\psi\}]L_j[\{\phi'\},\{\psi\}])=\text{independent of } \{\phi\} , \tag{84}$$

and so the second term on the right-hand side of Eq. (78) vanishes, giving in the multiquadratic case $S_{\rm eff}[\{\phi\},\{\psi\};\theta]=S[\{\phi\},\{\psi\}]$. In the application of the generalized algorithm to the Wilson lattice gauge theory given below, the second term on the right-hand side of Eq. (78) will be nonzero, but is arranged to be a higher-order correction in the continuum limit as compared with the original action S.

Let us now apply this algorithm to the Yang-Mills action

$$\beta S = \int d^4x \tfrac{1}{2}\mathrm{Tr}(F_{\mu\nu}F_{\mu\nu}),\quad F_{\mu\nu}=F^j_{\mu\nu}T^j,\quad \mathrm{Tr}(T^iT^j)=\tfrac{1}{2}\delta_{ij}\ , \tag{85}$$

with the field-strength $F_{\mu\nu}$ related to the potential A_μ by

$$A_\mu = A^j_\mu T^j,\quad F_{\mu\nu}=\partial_\mu A_\nu - \partial_\nu A_\mu + ig_0[A_\mu, A_\nu]\ . \tag{86}$$

To formulate a discrete version of Eq. (85), we set up a cubic lattice with unit cell of side a, and associate the potential variables with the centers of the links. Then for a plaquette in the x_μ-x_ν plane with center $(x_{c\mu}, x_{c\nu})$, as shown in Fig. 1, we have, for the field-strength component $F_{\mu\nu}$ at the center of the plaquette,

$$\begin{aligned} F_P \equiv F_{\mu\nu}(x_{c\mu},x_{c\nu}) = a^{-1}[&A_\nu(x_{c\mu}+\tfrac{1}{2}a, x_{c\nu}) - A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu}) - A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a) + A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a)] \\ &+ ig_0\tfrac{1}{2}[A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a), A_\nu(x_{c\mu}+\tfrac{1}{2}a,x_{c\nu})] \\ &+ ig_0\tfrac{1}{2}[A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a), A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu})] + O(a^2)\ , \end{aligned} \tag{87}$$

where the dependence on coordinates other than x_μ and x_ν is not shown explicitly. Summing over plaquettes, and noting that each plaquette is shared between two unit cells, we have for the discretized action

$$\beta S = \sum_P a^4 \mathrm{Tr}(F_P{}^2)\ . \tag{88}$$

Consider now an update in which the potential A_l on a single link l is changed. By Eq. (87), for each plaquette $P \supset l$, the field strength F_P is a linear functional of A_l, while for all other plaquettes the field strength has no dependence on A_l. Hence if we let $\{\phi\}$ be the set of potential components A^j_l, and $\{\psi\}$ be all other potential components, then in terms of these variables the action of Eq. (88) has precisely the form of Eqs. (76b) and (80). Moreover, if we choose any canonical gauge fixing (or if we do not gauge fix), then the integration measure has the form

$$d\mu = \prod \int dA^j_l \prod \int d\psi \tag{89}$$

required by Eq. (76a). Thus the conditions for validity of the algorithm of Eq. (82) are satisfied, and so an overrelaxed algorithm for the update $A_l \to A'_l$ is

$$\begin{aligned} W[A_l \to A'_l] &= \mathcal{N}\exp\left[-\sum_{P\supset l} a^4\mathrm{Tr}(\cosh\theta F'_P + \sinh\theta F_P)^2\right] = \mathcal{N}\exp\left[\frac{-1}{\omega(2-\omega)}\sum_{P\supset l} a^4 \mathrm{Tr}[F'_P-(1-\omega)F_P]^2\right]\ , \\ F'_P &= F_P\,|_{A_l\to A'_l}\ . \end{aligned} \tag{90}$$

Although Eq. (90) exactly satisfies detailed balance with respect to the discretized action of Eq. (88), it is not exactly gauge invariant, and this limits its usefulness in simulations where maintaining exact gauge invariance is important. To get a computationally useful algorithm, we must construct an analog of Eq. (90) within the framework of Wilson's lattice gauge theory.[14] Because the lattice gauge theory action is not a multiquadratic form, it is not possible to construct an overrelaxed algorithm which exactly satisfies detailed balance with respect to the Wilson lattice action.

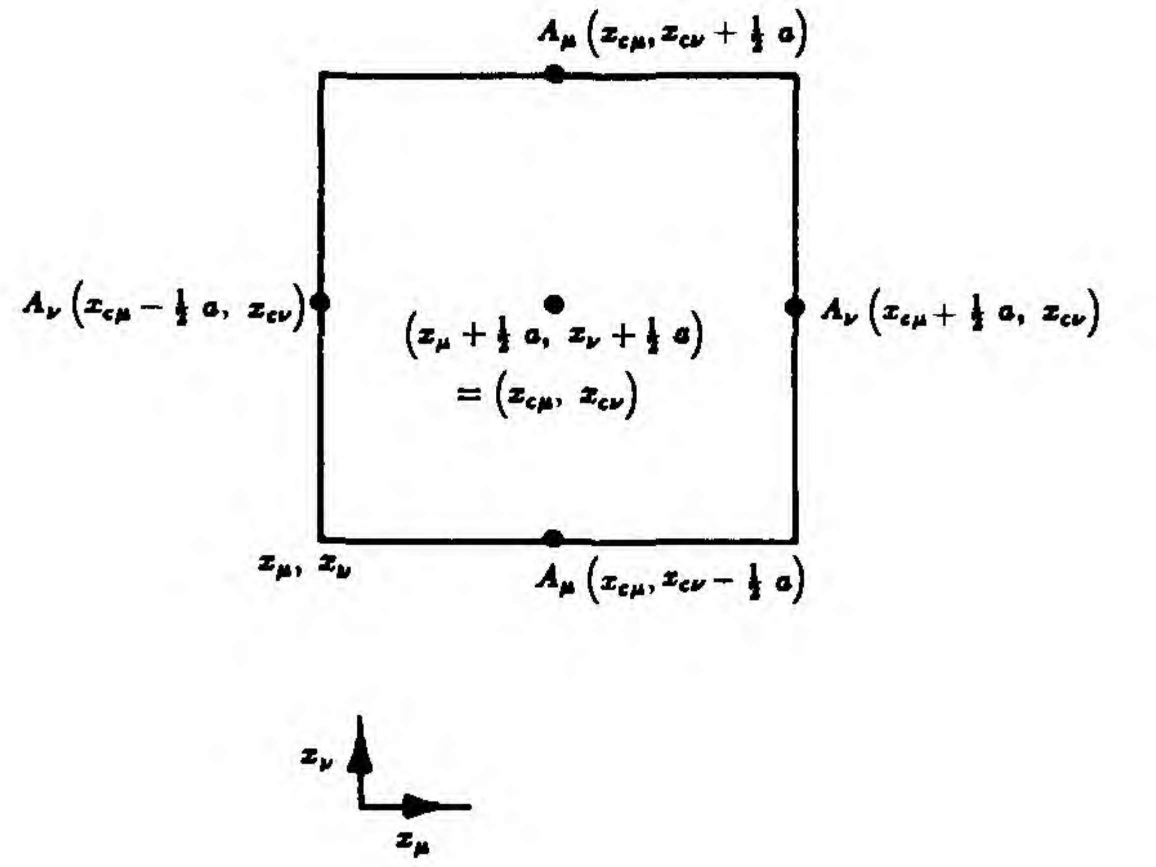


FIG. 1. Plaquette and potential variables in the x_μ-x_ν plane used to formulate Yang-Mills lattice field theory.

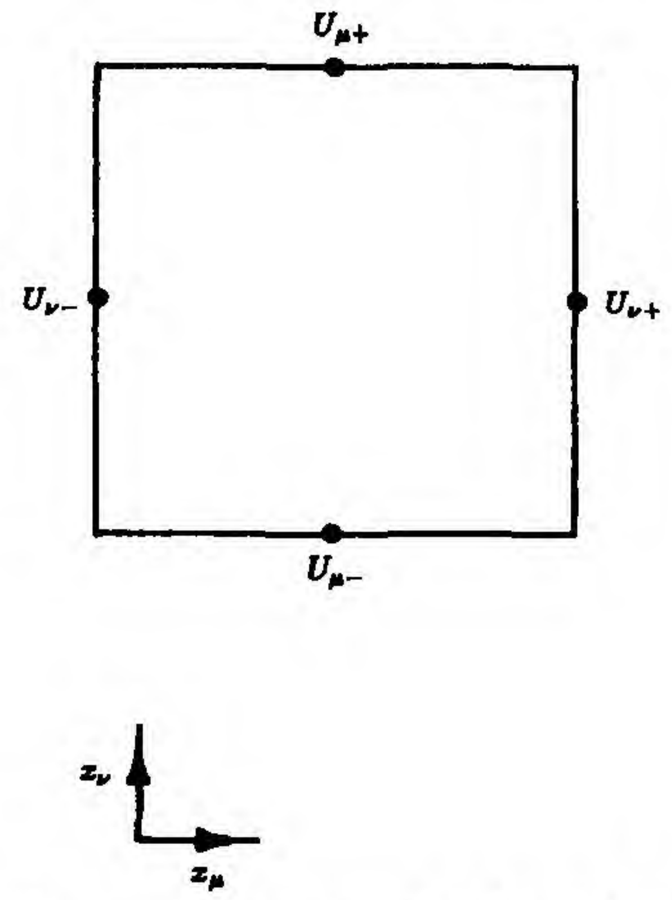


FIG. 2. Unitary matrices associated with the links of the plaquette of Fig. 1, which are used to formulate SU(n) lattice gauge theory.

However, this is a stronger requirement than is needed, since the lattice action is in any case only an order-a^2 approximation to the continuum action, and any member of the equivalence class of local, gauge-invariant lattice actions which differ from the Wilson action by relative order-a^2 terms in the continuum limit is equally suitable as a lattice action. We will show that it *is* possible to construct an exactly gauge-invariant lattice gauge theory transition probability by the procedure of Eqs. (76)–(78) above, which satisfies detailed balance with respect to an explicitly computable effective action differing from the Wilson action only by terms of relative order a^2 in the continuum limit.

To carry out this construction we rewrite Eq. (90) as

$$W[A_l \to A_l'] = \mathcal{N}\exp\Big[-\sinh\theta\cosh\theta \sum_{P\supset l} a^4\mathrm{Tr}(F_P'+F_P)^2 - (\cosh^2\theta - \sinh\theta\cosh\theta)\sum_{P\supset l} a^4\mathrm{Tr}(F_P')^2 - (\sinh^2\theta-\sinh\theta\cosh\theta)\sum_{P\supset l} a^4\mathrm{Tr}(F_P)^2\Big] \tag{91}$$

and look for lattice gauge theory realizations of $a^4\mathrm{Tr}(F_P)^2$, $a^4\mathrm{Tr}(F_{P'})^2$, and $a^4\mathrm{Tr}(F_{P'}+F_P)^2$. Consider the plaquette P drawn in Fig. 1; in Fig. 2 we have redrawn this plaquette with the links labeled by the SU(n) matrices to which they correspond in lattice gauge theory. Let us assume that the link potential being updated is $A_l = A_\mu(x_{c\mu}, x_{c\nu}-\frac{1}{2}a)$, or in terms of lattice gauge theory variables, $U_l = U_{\mu-}$. We define

$$U_P = U_{\mu-}U_{\nu+}U_{\mu+}U_{\nu-}\,,\quad U_P' = U_{\mu-}'U_{\nu+}U_{\mu+}U_{\nu-} = U_P\,|_{U_l\to U_l'}\,. \tag{92}$$

Then the lattice gauge theory analog of Eq. (91) is

$$\begin{aligned} W[U_l\to U_l'] &= \mathcal{N}\exp\Bigg[-\sinh\theta\cosh\theta \sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}(U_PU_P')\right] \\ &\quad -(\cosh^2\theta-\sinh\theta\cosh\theta)\sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}U_P'\right] \\ &\quad -(\sinh^2\theta-\sinh\theta\cosh\theta)\sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}U_P\right]\Bigg] \\ &= \mathcal{N}\exp\Bigg[\frac{1-\omega}{\omega(2-\omega)}\sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}(U_PU_P')\right] - \frac{1}{\omega}\sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}U_{P'}\right] \\ &\quad -\frac{1-\omega}{\omega}\sum_{P\supset U_l}\beta_0\left[1-\frac{1}{n}\mathrm{Re\,Tr}U_P\right]\Bigg]\,, \end{aligned} \tag{93}$$

with Re denoting the real part, with $\mathcal{N}$ fixed by the requirement

$$\int d[U_l']W[U_l\to U_l'] = 1\,, \tag{94}$$

and with the parameter β_0 fixed in terms of n and the bare coupling g_0 by the usual relation[15]

$$\frac{\beta_0 g_0{}^2}{2n} = 1\,. \tag{95}$$

We note that Eq. (93) is independent of the cyclic ordering of the link factors in U_P, as long as U_P' and U_P are ordered in the same way; in other words, by cyclic invariance of the trace we have

$$\begin{aligned} &\mathrm{Tr}[(U_{\mu-}U_{\nu+}U_{\mu+}U_{\nu-})(U_{\mu-}'U_{\nu+}U_{\mu+}U_{\nu-})] \\ &\quad = \mathrm{Tr}[(U_{\nu-}U_{\mu-}U_{\nu+}U_{\mu+})(U_{\nu-}U_{\mu-}'U_{\nu+}U_{\mu+})] \\ &\quad = \mathrm{Tr}[(U_{\mu+}U_{\nu-}U_{\mu-}U_{\nu+})(U_{\mu+}U_{\nu-}U_{\mu-}'U_{\nu+})] \\ &\quad = \cdots\,. \end{aligned} \tag{96}$$

The gauge invariance of Eq. (93) follows from the fact that since U_l and U_l' have the same behavior under gauge transformation, so do U_P and U_P':

$$U_P \to u_g U_P u_g{}^{-1}\,,\quad U_P' \to u_g U_P' u_g{}^{-1}\,, \tag{97}$$

and hence again by cyclic invariance of the trace the quantities $\mathrm{Tr}(U_PU_P')$, $\mathrm{Tr}U_P$, and $\mathrm{Tr}U_P'$ are exactly gauge invariant. Finally, since the U_l' dependence of Eq. (93) is of the form $\mathrm{Tr}(U_l'\tilde{U})$, with $\tilde{U}$ a linear combination of SU(n) matrices, in the case $n=2$ the efficient SU(2) algorithm of Creutz[16] can be applied to the generation of links U_l' distributed according to $W[U_l\to U_l']$.

The argument that Eq. (93) is an acceptable algorithm now runs as follows: Comparing with Eqs. (76)–(78), we see that Eq. (93) has precisely the form of the generalized algorithm, with $\{\phi\}$ corresponding to U_l, with $\{\psi\}$ corresponding to the other links in the plaquettes $P\supset U_l$, and with S_1 corresponding to those terms in the Wilson action involving plaquettes $P\supset U_l$. Hence Eq. (93) exactly satisfies detailed balance with respect to an

effective action, which differs from the Wilson action by a term proportional to $\ln(\mathcal{N}[U_l,\{\psi\};\theta]/\overline{\mathcal{N}}[\{\psi\},\theta])$, with $\overline{\mathcal{N}}$ the average of $\mathcal{N}[U_l,\{\psi\};\theta]$ over U_l. Since Eq. (93) only involves couplings of the link l to links in plaquettes P containing l, and since the entire construction is manifestly lattice gauge invariant, the effective action is local and lattice gauge invariant. Suppose that we can show that Eq. (93) differs from Eq. (91) by terms of relative order a^2 (absolute order a^6) in the continuum limit; then to leading order (absolute order a^4) the normalization factor $\mathcal{N}[U_l,\{\psi\};\theta]$ is independent of U_l, since by translation invariance $\mathcal{N}$ in Eq. (91) is independent of A_l. It then follows that $\ln(\mathcal{N}[U_l,\{\psi\};\theta]/\overline{\mathcal{N}}[\{\psi\};\theta])$ is of order a^6 in the continuum limit, and the equilibrium effective action for the algorithm of Eq. (93) is a member of the equivalence class of acceptable lattice actions.

To verify that in the continuum limit Eq. (93) reduces to Eq. (91) up to an error of order a^2, we start from the continuum limit of the individual link variables,

$$\begin{aligned} U_{\mu-} &= \exp[ig_0 a A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a)+O(a^3)] , \\ U_{\nu+} &= \exp[ig_0 a A_\nu(x_{c\mu}+\tfrac{1}{2}a,x_{c\nu})+O(a^3)] , \\ U_{\mu+} &= \exp[-ig_0 a A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a)+O(a^3)] , \\ U_{\nu-} &= \exp[-ig_0 a A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu})+O(a^3)] , \end{aligned} \tag{98}$$

with $\mathrm{Tr}O(a^3)=0$ since the U's are all SU(n) matrices and hence have unit determinant. For the products of adjacent links which appear in U_P, we have

$$\begin{aligned} &\exp[ig_0 a A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a)]\exp[ig_0 a A_\nu(x_{c\mu}+\tfrac{1}{2}a,x_{c\nu})]=e^{\Phi_+ +\delta_+} , \\ &\exp[-ig_0 a A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a)]\exp[-ig_0 a A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu})]=e^{\Phi_- +\delta_-} , \end{aligned} \tag{99}$$

with

$$\begin{aligned} \Phi_+ &= ig_0 a A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a)+ig_0 a A_\nu(x_{c\mu}+\tfrac{1}{2}a,x_{c\nu}) \\ &\quad -\tfrac{1}{2}g_0{}^2a^2[A_\mu(x_{c\mu},x_{c\nu}-\tfrac{1}{2}a),A_\nu(x_{c\mu}+\tfrac{1}{2}a,x_{c\nu})] , \\ \Phi_- &= -ig_0 a A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a)-ig_0 a A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu}) \\ &\quad -\tfrac{1}{2}g_0{}^2a^2[A_\mu(x_{c\mu},x_{c\nu}+\tfrac{1}{2}a),A_\nu(x_{c\mu}-\tfrac{1}{2}a,x_{c\nu})] . \end{aligned} \tag{100}$$

The errors δ_+ and δ_- satisfy $\delta_+=O(a^3)$, $\delta_-=O(a^3)$, $\mathrm{Tr}\delta_+=\mathrm{Tr}\delta_-=0$. Moreover, since $\delta_-=\delta_+(a\to -a)$ and $\Phi_-=\Phi_+(a\to -a)$, we have that $\delta_+ +\delta_-=O(a^4)$ and that the commutator $[\Phi_+,\Phi_-]$ is odd in a. Hence for the plaquette product U_P we have

$$\begin{aligned} U_P &= U_{\mu-}U_{\nu+}U_{\mu+}U_{\nu-} \\ &= \exp\{\Phi_+ +\Phi_- +\tfrac{1}{2}[\Phi_+,\Phi_-]+O(a^4)\} , \end{aligned} \tag{101}$$

with $\frac{1}{2}[\Phi_+,\Phi_-]=O(a^3)$, with $\mathrm{Tr}O(a^4)=0$ and [referring to Eq. (87)] with

$$\Phi_+ +\Phi_- = ig_0 a^2 F_P . \tag{102}$$

For a general altered set of potentials A' we have

$$\begin{aligned} U'_P &= U'_{\mu-}U'_{\nu+}U'_{\mu+}U'_{\nu-} \\ &= \exp\{\Phi'_+ +\Phi'_- +\tfrac{1}{2}[\Phi'_+,\Phi'_-]+O(a^4)\} , \end{aligned} \tag{103}$$

again with $\frac{1}{2}[\Phi'_+,\Phi'_-]=O(a^3)$, with $\mathrm{Tr}O(a^4)=0$, and with

$$\Phi'_+ +\Phi'_- = ig_0 a^2 F'_P . \tag{104}$$

From these equations we find

$$\begin{aligned} \mathrm{Tr}U_P &= n-\tfrac{1}{2}g_0{}^2a^4\mathrm{Tr}(F_P)^2+O(a^6) , \\ \mathrm{Tr}U'_P &= n-\tfrac{1}{2}g_0{}^2a^4\mathrm{Tr}(F'_P)^2+O(a^6) , \\ \mathrm{Tr}(U_PU'_P) &= n-\tfrac{1}{2}g_0{}^2a^4\mathrm{Tr}(F_P+F'_P)^2+\Delta+O(a^6) , \\ \Delta &= \mathrm{Tr}\{(\Phi_+ +\Phi_- +\Phi'_+ +\Phi'_-)(\tfrac{1}{2}[\Phi_+,\Phi_-] \\ &\qquad +\tfrac{1}{2}[\Phi'_+,\Phi'_-])\} , \end{aligned} \tag{105}$$

which when $\Delta=0$ can be combined with Eqs. (95) and (93) to give Eq. (91). (Because of the identity $\mathrm{Tr}(\alpha[\alpha,\gamma])=0$, terms analogous to Δ do not appear in $\mathrm{Tr}U_P$ and $\mathrm{Tr}U'_P$.) Since the error term Δ is potentially of order a^5, to complete the derivation we must show that $\Delta=0$. Now repeated use of the identity

$$\mathrm{Tr}(\alpha[\beta,\gamma])=\mathrm{Tr}([\gamma,\alpha]\beta) , \tag{106}$$

which follows from cyclic invariance of the trace, shows that Δ can be reduced to the form

$$\Delta=\tfrac{1}{2}\mathrm{Tr}\{[\Phi'_+ -\Phi_+,\Phi'_- -\Phi_-](\Phi_+ +\Phi_-)\} . \tag{107}$$

In general $\Delta\neq 0$, but for the special case in which A' differs from A by the change of only the single link variable $A_\mu(x_{c\mu},x_{c\nu}-\frac{1}{2}a)$ or equivalently $U_{\mu-}$, we have $\Phi'_-=\Phi_-$ and Δ vanishes. Hence we have verified that Eq. (93) is a suitable overrelaxed algorithm for lattice gauge theory, for the case in which a single link at a time is updated.

To conclude this section, let us compare the small-ω continuum limit of the overrelaxation algorithm of Eq. (93) with the continuum limit of the lattice Langevin algorithm of Batrouni *et al.*; according to our analysis of Sec. I, these should correspond. Taking the continuum limit of the overrelaxation algorithm from Eq. (90), we have

$$\mathrm{Tr}(F_P'-F_P+\omega F_P)^2=\mathrm{Tr}\left[\frac{1}{a}(A_l'-A_l)+\tfrac{1}{2}ig_0[(A_l'-A_l),A_{\text{adjacent}}]+\omega F_P\right]^2, \tag{108a}$$

with A_{adjacent} the potential on the leg of the plaquette adjacent to and following A_l. Referring to Eq. (98), we recall that in the continuum limit g_0aA is the effective expansion parameter; approximating Eq. (108a) to leading-order accuracy in this expansion gives

$$\mathrm{Tr}(F_P'-F_P+\omega F_P)^2\approx\mathrm{Tr}\left[\frac{1}{a}(A_l'-A_l)+\omega F_P\right]^2=\frac{1}{2}\frac{1}{a^2}(A_l'^j-A_l^j)^2+\frac{1}{a}(A_l'^j-A_l^j)\omega F_P^j+A_l'\text{-independent}. \tag{108b}$$

In four dimensions there are six plaquettes P containing l, and so

$$\sum_{P\supset l}\mathrm{Tr}(F_P'-F_P+\omega F_P)^2\approx\frac{3}{a^2}(A_l'^j-A_l^j)^2+\frac{1}{a}(A_l'^j-A_l^j)\omega\sum_{P\supset l}F_P^j+A_l'\text{-independent}$$
$$=3\left[\frac{1}{a}(A_l'^j-A_l^j)+\frac{\omega}{6}\sum_{P\supset l}F_P^j\right]^2+A_l'\text{-independent}. \tag{109}$$

Substituting Eq. (109) into Eq. (90), dropping A_l'-independent terms and approximating $2-\omega\approx 2$, we have

$$W[A_l\to A_l']\approx\mathcal{N}\exp\left[-\tfrac{1}{4}a^4\frac{6}{\omega}\left[\frac{1}{a}(A_l'^j-A_l^j)+\frac{\omega}{6}\sum_{P\supset l}F_P^j\right]^2\right], \tag{110}$$

which can be rewritten as the stochastic difference equation

$$\frac{1}{a}(A_l'^j-A_l^j)\approx-\frac{\omega}{6}\sum_{P\supset l}F_P^j-\left[\frac{\omega}{6}\right]^{1/2}\frac{1}{a^2}\eta_j. \tag{111}$$

Now the lattice Langevin algorithm of Batrouni *et al.*, in the notation used above, takes the form

$$U_l'=e^{-F_l}U_l,$$
$$F_l=iT^j\left[\bar\epsilon\sum_{P\supset l}\left[\frac{-i\beta_0}{2n}\right]\mathrm{Tr}[T^j(U_P-U_P^\dagger)]+\bar\epsilon^{1/2}\eta_j\right]. \tag{112}$$

Substituting

$$U_l=e^{ig_0aA_l},\quad U_P=e^{ig_0a^2F_P} \tag{113}$$

into Eq. (112) and working to leading order in the expansion in powers of g_0aA, we get, for the continuum limit,

$$ig_0a(A_l'^j-A_l^j)T^j\approx-iT^j\left[\bar\epsilon\left[\frac{-i\beta_0}{2n}\right]\sum_{P\supset l}ig_0a^2F_P^j+(\bar\epsilon)^{1/2}\eta_j\right], \tag{114}$$

which on substituting $\beta_0g_0^2/(2n)=1$ and factoring away the generators T^j becomes

$$\frac{1}{a}(A_l'^j-A_l^j)=-\frac{\bar\epsilon}{g_0^2}\sum_{P\supset l}F_P^j-\left[\frac{\bar\epsilon}{g_0^2}\right]^{1/2}\frac{1}{a^2}\eta_j. \tag{115}$$

Equations (111) and (115) have precisely the same structure, and give the identification

$$\omega=\frac{6}{g_0^2}\bar\epsilon, \tag{116}$$

again showing that the Langevin approach corresponds to the small-ω limit of the overrelaxation algorithm.

IV. DISCUSSION

In closing I comment briefly on the comparison between the acceleration strategy pursued above and that proposed by Batrouni *et al.*[9] Let us adopt as the "figure of merit" for an acceleration scheme the ratio of its inverse correlation time λ to that for an $\omega=1$ Jacobi iteration. As we have seen in Sec. II, for an optimally overrelaxed Gauss-Seidel iteration, the figure of merit is then N, the length of a side of the lattice in lattice units. By contrast, Batrouni *et al.* employ a Langevin method based on the Jacobi algorithm, and propose a method of Fourier acceleration in which the Langevin step size is taken to have a momentum dependence which compensates the critical slowing down at long wavelengths. In principle, their method can yield (up to logarithms) an inverse correlation time of $\lambda\sim\bar\epsilon a^{-1}$, with a the lattice spacing and $\bar\epsilon$ the small parameter which governs the Langevin step size. Thus, recalling Eq. (71), for the method of Batrouni *et al.*, the figure of merit can be as large as

$$\frac{\bar\epsilon a^{-1}}{ak_1^2}\sim\bar\epsilon(L/a)^2=\bar\epsilon N^2. \tag{117}$$

For lattices of moderate size, where $\bar\epsilon N\sim 1$, the overrelaxation method should be competitive with Fourier acceleration, but for very large lattices the Fourier method wins out, irrespective of the step size $\bar\epsilon$. Clearly, an optimal algorithm would combine the advantages of both, by permitting a step size of unity, as in the overrelaxed Gauss-Seidel approach, while replacing the factor $k_1\sim L^{-1}$ in Eq. (69) by a wave number of order a^{-1}. One possible way to try to achieve an improved algorithm is to combine overrelaxation with a mesh-doubling

lattice refinement scheme, as is done in the case of the minimization problem by the "hyper-overrelaxation" algorithm of Press[17] or the mesh-refinement-interpolation scheme of Adler and Piran.[7] A closely related approach is the "multigrid" Monte Carlo method advocated by Goodman and Sokal.[10] I hope to pursue these issues in future work.

Note added

In Sec. II we determined an optimum value of ω—let us call it ω_b–defined as the value of ω which minimizes the correlation time τ. By definition, this value of ω maximizes the asymptotic rate of decay of the correlation between two lattice configurations, as the "time" separation $\Delta T=\Delta Ma$ between the two configurations becomes infinite (a=lattice spacing, ΔM=number of iterations separating the two configurations). However, in an actual Monte Carlo calculation this asymptotic decay rate is not the quantity which directly governs errors. What one does in a Monte Carlo calculation is to perform some total number M of iterations, but to only take every mth iterate as a member of the ensemble of configurations used for measurements, where $m\sim\tau/a$. Taking more configurations than this increases the amount of effort spent in measurement without improving the statistics, since the additional configurations are not statistically independent, while taking fewer configurations than this needlessly dilutes the statistics. Hence the quantity to be optimized is the absolute correlation between two configurations separated by m iterations, not the asymptotic rate of correlation decay.

This optimization problem also arises in the overrelaxation solution of differential equations, and the solution is as follows: For the iterations $i=0,1,\ldots$ one uses overrelaxation with a sequence of relaxation parameters ω_i with $\omega_0=1$ and with $\omega_i\to\omega_b$ for large i. In the case of iterations based on "odd/even" or "checkerboard" ordering, as opposed to the "typewriter" ordering used in Sec. II, the optimum ω_i's can be computed explicitly in terms of i and ω_b using Chebyshev polynomials. In the Monte Carlo application, one would use a "sawtooth" pattern of ω's, returning ω to 1 for the initial iteration after each configuration selected for measurement, and then stepping through the first m members of the Chebyshev or other optimal sequence. Taking $\omega\equiv\omega_b$ for all iterations can actually make the correlations *worse* after a finite number of iterations than simply using $\omega=1$, while the simple expedient of taking $\omega_0=1$ and $\omega_{i>1}=\omega_b$ already guarantees monotonically decreasing correlations. For a brief and lucid discussion of these issues see Hockney and Eastwood,[18] while for a detailed theoretical analysis see Vargas.[19] A simple, explicit, "checkerboard" iteration version of the calculation of Sec. II has recently been given by Neuberger,[20] and the Chebyshev method is directly applicable to Neuberger's scheme.

ACKNOWLEDGMENTS

I wish to thank F. Brown, M. Creutz, and C. Whitmer for stimulating conversations about lattice algorithms, and for informing me before publication of their numerical work suggesting that overrelaxation improves the correlation time. I also wish to thank G. P. Lepage for helpful conversations about the Langevin algorithm and the work of the Cornell group of Batrouni *et al.* I am grateful to M. Rassetti and T. Regge for the hospitality of the Institute for Scientific Interchange in Torino, and to E. Predazzi for the hospitality of the University of Torino and the use of its physics library. Partial support for this publication was provided by the U.S. Department of Energy under Grant No. DE-AC02-76ERO2220.

APPENDIX: EVALUATION OF THE INFINITE-SPACE GREEN'S FUNCTION

We evaluate here the time dependence of the Gauss-Seidel Green's function of Eq. (60), in the infinite-space limit in which the mode functions are

$$\psi_m(x)=\frac{1}{(2\pi)^{d/2}}e^{i\mathbf{k}\cdot\mathbf{x}},\quad \mathbf{k}\cdot\mathbf{x}=\sum_{\mu=1}^{d}k_\mu x_\mu\,. \tag{A1}$$

Substituting into Eq. (60), and noting that because of translation invariance there is no loss of generality in setting $x'=t'=0$, we have

$$\begin{aligned}
&G(x,0;t,0)=\frac{4}{d}\frac{2}{\omega}\frac{a^{d-1}}{(2\pi)^d}\int d^dk\, e^{i\mathbf{k}\cdot\mathbf{x}}\frac{e^{-pt-p\mathbf{x}\cdot\mathbf{n}/2}-e^{-qt-q\mathbf{x}\cdot\mathbf{n}/2}}{q-p}\theta(t+\tfrac{1}{2}\mathbf{x}\cdot\mathbf{n})\,,\\
&\mathbf{n}=(1,1,\ldots,1),\quad \mathbf{n}\cdot\mathbf{x}=\sum_{\mu=1}^{d}x_\mu,\quad p=2[C-(C^2-k^2/d)^{1/2}]\,,\\
&q=2[C+(C^2-k^2/d)^{1/2}],\quad k^2=\sum_{\mu=1}^{d}k_\mu{}^2\,.
\end{aligned}\tag{A2}$$

We wish to evaluate the Fourier transform $g(l,t)$ defined by

$$G(x,0;t,0)=\frac{4}{d}\frac{2}{\omega}\frac{a^{d-1}}{(2\pi)^d}\int d^dl\, e^{il\cdot x}g(l,t)\,. \tag{A3}$$

Taking the inverse Fourier transform of Eq. (A2), the x and k integrations in the $d-1$ directions perpendicular to $\mathbf{n}$ can be done immediately, leaving [with $\mathbf{n}=d^{1/2}\hat{\mathbf{n}}$, $x=\mathbf{x}\cdot\hat{\mathbf{n}}$, $l=l\cdot\hat{\mathbf{n}}$, $k=\mathbf{k}\cdot\hat{\mathbf{n}}$, $(l^{\perp})^2=l^2-l^2$]

$$g(l,l^{\perp},t)=\int \frac{dx}{2\pi} e^{-ilx} \int dk\, e^{ikx} \frac{e^{-pl-pd^{1/2}x/2}-e^{-ql-qd^{1/2}x/2}}{q-p}\theta(t+\tfrac{1}{2}d^{1/2}x) ,$$
$$\left.\begin{matrix} p \\ q \end{matrix}\right\} =2(C\mp\{C^2-[k^2+(l^{\perp})^2]/d\}^{1/2}) . \tag{A4}$$

To do the x integration, we make the change of variables

$$t+\tfrac{1}{2}d^{1/2}x=u , \tag{A5}$$

giving

$$g(l,l^{\perp},t)=\frac{1}{\pi d^{1/2}}\int_{-\infty}^{\infty} dk\, e^{-i(k-l)2t/d^{1/2}} I_u ,$$
$$I_u=\int_0^{\infty} du\, e^{i(k-l)2u/d^{1/2}} \frac{e^{-pu}-e^{-qu}}{q-p}=\frac{1}{q-p}\left[\frac{1}{p-i(k-l)2/d^{1/2}}-\frac{1}{q-i(k-l)2/d^{1/2}}\right]$$
$$=\frac{1}{pq-i(k-l)2(p+q)/d^{1/2}-4(k-l)^2/d} . \tag{A6}$$

Substituting p and q from Eq. (A4), and setting $k-l=w$, we are left with the single integral

$$g(l,l^{\perp},t)=\frac{1}{\pi d^{1/2}}\int_{-\infty}^{\infty} dw\, e^{-iw2t/d^{1/2}} \frac{1}{4[l^2+(l^{\perp})^2+2lw]/d-iw8C/d^{1/2}} . \tag{A7}$$

Since the denominator has a single zero in the lower half of the w complex plane, for $t<0$ we can close the contour up to get $g=0$, while for $t>0$ we can close the contour down to get the answer quoted in Eq. (75) of the text.

---

*Permanent address.

[1]See K. Binder, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 5b; M. Creutz, *Quarks, Gluons and Lattices* (Cambridge University Press, Cambridge, England, 1983), Chap. 18.

[2]S. L. Adler, Phys. Rev. D **23**, 2901 (1981). This earlier paper contains an inadvertent normalization error which is corrected in Eqs. (9) and (11) below.

[3]S. L. Adler and T. Piran, in *High-Energy Physics—1980*, proceedings of the XXth International Conference, Madison, Wisconsin, edited by L. Durand and L. G. Pondrom (AIP Conf. Proc. No. 68) (AIP, New York, 1981).

[4]A connection between underrelaxation and the Langevin approach has been independently noted, in a qualitative way, by M. Creutz, Phys. Rev. D **36**, 515 (1987).

[5]C. Whitmer, Phys. Rev. D **29**, 306 (1984); thesis, Princeton University, 1984.

[6]F. R. Brown and T. J. Woch, Phys. Rev. Lett. **26**, 2394 (1987).

[7]This is in general the case for the minimization problem; see S. L. Adler and T. Piran, Rev. Mod. Phys. **56**, 1 (1984). However, the *equality* of the thermalization and the correlation exponents found below is a feature only of the quadratic action, or free field, case.

[8]Although the choice of boundary conditions should not matter for sufficiently large lattices, the use of periodic boundary conditions leads to noncausal "wave propagation" in the continuum analog of the Gauss-Seidel iteration, and hence does not permit the use of the causality argument of Eqs. (61)–(63) below to solve the $t=0$ boundary condition.

[9]G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).

[10]J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986); work in preparation; A. D. Sokal (private communication).

[11]P. R. Garabedian, Math. Tables Aids Comp. **10**, 183 (1956); see also W. F. Ames, *Numerical Methods for Partial Differential Equations* (Academic, New York, 1977), Chap. 3.

[12]A hyperbolic equation normally requires two initial conditions on an initial-value surface which is not a characteristic. However, the surface $t=0$ is a characteristic of Eq. (40) and so this general rule does not apply. In fact, since $dC+\sum_{\mu=1}^{d}\partial/\partial x_\mu$ is an invertible operator, the time development of G and $\bar{\phi}$ can be integrated forward from the values of G and $\bar{\phi}$ for all x at $t=0$.

[13]An alternative argument is to use the discrete form of the iteration for G given in Eqs. (28a) and (29) to infer that $G_{i,i'}^{n,n'}=0$ unless $n\geq n'$ and $n+i_1+\cdots+i_\mu\geq n'+i_1'+\cdots+i_\mu'$. The continuum limit of these inequalities bounding the region of support is $t-t'\geq 0$ and $t-t'+\sum_{\mu=1}^{d}(x_\mu-x_\mu')\geq 0$. Adding the two inequalities we learn that G has support only in $s-s'\geq 0$, and so $G_m^<=0$ is the correct boundary condition.

[14]K. Wilson, Phys. Rev. D **10**, 2445 (1974).

[15]Creutz, *Quarks, Gluons and Lattices* (Ref. 1), p. 36.

[16]Creutz, *Quarks, Gluons and Lattices* (Ref. 1), pp. 132 and 133.

[17]W. Press, Harvard-Smithsonian Center for Astrophysics Report No. 971, 1978 (unpublished).

[18]R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981), pp. 174–181.

[19]R. S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1962), Chap. 5.

[20]H. Neuberger, Phys. Rev. Lett. **59**, 1877 (1987).

VOLUME 60, NUMBER 13 PHYSICAL REVIEW LETTERS 28 MARCH 1988

Stochastic Algorithm Corresponding to a General Linear Iterative Process

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 16 November 1987)

Let $u'=Mu+Nf$ be a general linear iterative process for solving the system $Lu=f$, with $L=L^{\mathrm{T}}$ and with $1=M+NL$. Provided that $\Gamma\equiv\frac{1}{2}(L^{-1}-ML^{-1}M^{\mathrm{T}})^{-1}$ is a positive-definite matrix, it is shown that one can explicitly construct a corresponding stochastic algorithm which satisfies the homogeneous-state condition with respect to the probability distribution $\exp(-\beta S)$, where $S=\frac{1}{2}u^{\mathrm{T}}Lu-f^{\mathrm{T}}u$. When $M^{\mathrm{T}}L=LM$, the algorithm also satisfies the detailed-balance condition.

PACS numbers: 11.15.Ha, 02.70.+d

Over the last few years there has been considerable interest in the idea of the construction of accelerated Monte Carlo algorithms by analogy with acceleration schemes for solving deterministic systems of equations. For the physically interesting case of quadratic or multiquadratic actions, I showed[1] a while ago that one can construct an exact stochastic analog of the classical successive overrelaxation method, and this observation has had a number of applications.[2] Recently, several authors have proposed applying more powerful methods for deterministic systems, such as the multigrid iteration[3] or fast Fourier transform direct-solution techniques,[4] to the problem of accelerating Monte Carlo calculations. In this Letter I show, in the case of a quadratic action, that all of these proposals are closely related, and in fact are special cases of a theorem relating the most general linear iterative process to a corresponding stochastic algorithm.

Theorem.—Let $L=L^{\mathrm{T}}$ be a real, symmetric $l\times l$ matrix and f a real l-dimensional vector, from which we construct the action

$$S(u)=\tfrac{1}{2}u^{\mathrm{T}}Lu-f^{\mathrm{T}}u, \tag{1}$$

with variational equations

$$Lu=f. \tag{2}$$

Let us consider the general linear iteration for solving Eq. (2),

$$u'=Mu+Nf, \tag{3}$$

where M and N are a splitting of L defined by

$$1=M+NL. \tag{4}$$

Then provided that Γ as defined below is a positive definite matrix, corresponding to Eqs. (3) and (4) we can construct a stochastic process with normalized transition probability

$$P(u\rightarrow u')=(\beta/\pi)^{l/2}(\det\Gamma)^{1/2}\exp[-(u'-Mu-Nf)^{\mathrm{T}}\beta\Gamma(u'-Mu-Nf)],$$
$$\Gamma=\tfrac{1}{2}(L^{-1}-ML^{-1}M^{\mathrm{T}})^{-1}=\Gamma^{\mathrm{T}}, \tag{5}$$

which satisfies the homogeneous-state condition

$$\int du\, e^{-\beta S(u)}P(u\rightarrow u')=e^{-\beta S(u')}. \tag{6}$$

The transition probability of Eq. (5) satisfies the stronger detailed-balance condition [which is sufficient but not necessary for Eq. (6)]

$$e^{-\beta S(u)}P(u\rightarrow u')=e^{-\beta S(u')}P(u'\rightarrow u) \tag{7}$$

if and only if $M^{\mathrm{T}}L=LM$, or equivalently $N=N^{\mathrm{T}}$.

Remarks.—(1) The overrelaxation[5] and multigrid[6] methods are special cases of the iteration of Eq. (4) with $M\neq 0$ and with spectral radius $\rho(M)<1$, while direct solution methods correspond to taking $M=0$, $N=L^{-1}$. The matrix Γ generalizes the temperature rescaling found, in the case of stochastic overrelaxation, in Ref. 1. The particular significance of stochastic overrelaxation, within the general framework of the theorem, is that one can show that it is the unique limiting case of the algorithm of Eq. (5) in which the node variables are updated one at a time.

(2) Equation (5) can be equivalently written as a stochastic difference equation. Letting O be the real, orthogonal matrix which diagonalizes Γ,

$$O^{\mathrm{T}}\Gamma O=\gamma=\text{diagonal}, \tag{8}$$

we have

$$u'=Mu+Nf+(2\beta^{1/2})^{-1}O\eta, \tag{9}$$

with η_i Gaussian noise variables normalized according to

$$\langle\eta_i\eta_j\rangle=2(\gamma^{-1})_{ij}. \tag{10}$$

(3) The positivity condition on Γ can be reexpressed in

a number of equivalent forms. Rewriting Γ as

$$\Gamma = \tfrac{1}{2}L^{1/2}(1-S^{\mathrm{T}}S)^{-1}L^{1/2}, \qquad S = L^{-1/2}M^{\mathrm{T}}L^{1/2}, \tag{11}$$

we see that Γ is positive definite if and only if $S^{\mathrm{T}}S$ has no eigenvalues larger than 1. This requires

$$\|S\|^2 = \rho(S^{\mathrm{T}}S) = \rho(LML^{-1}M^{\mathrm{T}}) \le 1, \tag{12}$$

with $\rho(\ )$ (as above) the spectral radius and $\|\ \|$ the spectral norm.[7] Convergence of the iteration of Eq. (3) requires

$$\rho(M) = \rho(M^{\mathrm{T}}) = \rho(S) = \rho(S^{\mathrm{T}}) < 1, \tag{13}$$

but since $\rho(S) < \|S\|$ for a non-Hermitian S, Eq. (13) does not imply Eq. (12).

(4) Even when positivity of Γ can be demonstrated, the practical implementation of the algorithm depends on the ease of constructing the matrix O of Eq. (8) which diagonalizes Γ, or equivalently, of find the matrix $T = O\gamma^{-1/2}$ which factorizes Γ^{-1} according to

$$\Gamma^{-1} = TT^{\mathrm{T}}, \tag{14}$$

so that the generalized noise can be constructed as

$$O\eta = T\bar{\eta}, \quad \langle \bar{\eta}_i \bar{\eta}_j \rangle = 2\delta_{ij}. \tag{15}$$

Conversely, in cases (such as those discussed in Refs. 1-4) in which a stochastic differential equation has been constructed with Eq. (1) as its equilibrium action, the factorization of Eq. (14) is explicitly established and this guarantees the positivity of Γ.

(5) Although the theorem applies only to quadratic actions, it is directly relevant to the Yang-Mills action and other multiquadratic[8] interacting theories. More generally, most practical methods for dealing with nonlinear problems are based on generalizations from linear methods,[9] and so our theorem can be expected to have implications for the development of Monte Carlo algorithms for nonlinear problems.

Proof of the Theorem.— We assume in intermediate steps of the proof of existence of M^{-1}, but since the final results only involve M, the case where M is not invertible can be obtained by continuity as a limit from the case where M^{-1} exists. Making the shifts $u = v + L^{-1}f$, $u' = v' + L^{-1}f$ in Eqs. (6) and (7) permits one to factor away the explicit f dependence; hence it suffices to prove the theorem in the case $f = 0$. Doing the integral in Eq. (6) by completing the square gives us

$$(\det\Gamma/\det\tilde{\Gamma})^{1/2} e^{-\beta\tilde{S}(u')},$$

$$\tilde{S}(u') = u'^{\mathrm{T}}\Gamma u' - (M^{\mathrm{T}}\Gamma u')^{\mathrm{T}}\tilde{\Gamma}^{-1}M^{\mathrm{T}}\Gamma u', \tag{16}$$

$$\tilde{\Gamma} \equiv \tfrac{1}{2}L + M^{\mathrm{T}}\Gamma M.$$

Equating the left- and right-hand sides of Eq. (6), we get two conditions: (i) overall normalization,

$$\det\Gamma = \det\tilde{\Gamma}; \tag{17}$$

(ii) $u'^{\mathrm{T}} \cdots u'$ term in exponent,

$$\Gamma - \tfrac{1}{2}L = \Gamma M\tilde{\Gamma}^{-1}M^{\mathrm{T}}\Gamma. \tag{18}$$

From the definition of $\tilde{\Gamma}$ we have

$$\tilde{\Gamma} - \tfrac{1}{2}L = M^{\mathrm{T}}\Gamma M; \tag{19}$$

if we multiply Eq. (18) by M on the right and use Eq. (19) to eliminate $M^{\mathrm{T}}\Gamma M$ we get

$$\Gamma M - \tfrac{1}{2}LM = \Gamma M\tilde{\Gamma}^{-1}(\tilde{\Gamma} - \tfrac{1}{2}L) = \Gamma M - \tfrac{1}{2}\Gamma M\tilde{\Gamma}^{-1}L. \tag{20}$$

This implies

$$\Gamma = LML^{-1}\tilde{\Gamma}M^{-1}, \tag{21}$$

the determinant of which gives Eq. (17). The elimination of $\tilde{\Gamma}$ by use of Eq. (19) gives

$$\Gamma = LML^{-1}(\tfrac{1}{2}LM^{-1} + M^{\mathrm{T}}\Gamma) = \tfrac{1}{2}L + LML^{-1}M^{\mathrm{T}}\Gamma, \tag{22}$$

which can be immediately solved to give the result Γ quoted in Eq. (5). For $\tilde{\Gamma}$, we then get

$$\tilde{\Gamma} = \tfrac{1}{2}(L^{-1} - L^{-1}M^{\mathrm{T}}LML^{-1})^{-1}. \tag{23}$$

If in addition to the stationary-state condition we impose the detailed-balance condition on P, we get the additional constraints: (iii) $u^{\mathrm{T}} \cdots u$ term in exponent,

$$\Gamma = \tilde{\Gamma}; \tag{24}$$

(iv) $u^{\mathrm{T}} \cdots u'$ term in exponent,

$$\Gamma M = M^{\mathrm{T}}\Gamma. \tag{25}$$

Substituting Eqs. (24) and (25) into Eq. (16) for $\tilde{\Gamma}$, we get

$$\Gamma = \tfrac{1}{2}L + \Gamma M^2, \tag{26}$$

which implies

$$\Gamma = \tfrac{1}{2}L(1-M^2)^{-1} = \tfrac{1}{2}(L^{-1} - M^2L^{-1})^{-1}. \tag{27}$$

The comparison of Eq. (27) with Eq. (5) then implies $L^{-1}M^{\mathrm{T}} = ML^{-1}$, or equivalently

$$M^{\mathrm{T}}L = LM. \tag{28}$$

Since by Eq. (4), $N = L^{-1} - ML^{-1}$, Eq. (28) is also equivalent to

$$N = N^{\mathrm{T}}. \tag{29}$$

This work was supported by the U.S. Department of Energy under Grant No. DE AC02-76ERO2220.

[1]S. L. Adler, Phys. Rev. D **23**, 2901 (1981).
[2]C. Whitmer, Phys. Rev. D **29**, 306 (1984); F. R. Brown and T. J. Woch, Phys. Rev. Lett. **58**, 2394 (1987); M. Creutz, Phys. Rev. D **36**, 515 (1987); S. L. Adler, Phys. Rev. D **37**, 458 (1988); H. Neuberger,Phys. Rev. Lett. **59**, 1877 (1987).
[3]J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986), and unpublished; R. C. Brower, K. J. M. Moriarty, E. Meyers, and C. Rebbi, in *Proceedings of the Third Copper Mountain Conference on Multigrid Methods, Copper Mountain, Colorado, 1987,* edited by S. McCormick (Dekker, New York, 1988).
[4]G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).
[5]See, e.g., R. S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, 1962).
[6]See, e.g., W. Hackbusch, *Multi-Grid Methods and Applications* (Springer-Verlag, Berlin, Heidelberg, 1985).
[7]See R. S. Varga, Ref. 5, pp. 9–11, for the definition of the spectral norm $\| \;\|$ and theorems relating it to the spectral radius. In Eqs. (12) and (13) we have used the fact that the eigenvalues of a matrix, and hence the spectral radius, are unchanged by a similarity transformation.
[8]S. L. Adler and T. Piran, in *High Energy Physics—1980,* edited by L. Durand and L. G. Pondrum, AIP Conference Proceedings No. 68, Particles and Fields Subseries No. 22 (American Institute of Physics, New York, 1981), Vol. 2, p. 958; Whitmer, Ref. 2; Adler, Ref. 2.
[9]S. L. Adler and T. Piran, Rev. Mod. Phys. **56**, 1 (1984), Sec. III F.

VOLUME 62, NUMBER 2 PHYSICAL REVIEW LETTERS 9 JANUARY 1989

Study of an Overrelaxation Method for Gauge Theories

Stephen L. Adler
The Institute for Advanced Study, Princeton, New Jersey 08540

Gyan V. Bhanot
Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306
(Received 29 August 1988)

We study the overrelaxation algorithm developed by one of us (S.A.) for the SU(2) gauge theory in four dimensions. We find improvement in the decorrelation time for the plaquette by a factor of 2 to 3. However, our results suggest that the Monte Carlo application of overrelaxation behaves quite differently from the classical differential-equation analog. In particular, the optimum value of the overrelaxation parameter ω may depend on what order parameter one measures. In addition, there are order parameters that do not appear to improve by overrelaxation.

PACS numbers: 11.15.Ha

Statistical systems close to criticality are difficult to simulate. This is because most simulation algorithms use local updating which results in a highly correlated sequence of configurations. Information about the update spreads amongst the variables via a diffusion process and, hence, to decorrelate the system over length scales of the order of the correlation length ξ takes on the order of

$$\tau \sim \xi^2 \tag{1}$$

steps. τ in Eq. (1) is usually called the decorrelation time.

Several methods to decrease τ have been suggested. A very popular recent method is to use overrelaxed algorithms.[1] Here, in analogy with methods invented for differential equations, an overrelaxation parameter $\omega \in [1,2)$ is introduced into the simulation. The Markov process is defined so that for any ω one obtains the correct asymptotic distribution. $\omega=1$ corresponds to the usual (nonoverrelaxed) updating. $\omega>1$ gives overrelaxation. In this Letter, we test a specific overrelaxation algorithm for pure-gauge-theory simulations[2] suggested by one of us (S.A.). It is simplest to implement for the SU(2) gauge theory and so we restrict ourselves to this case in $d=4$ dimensions with periodic boundary conditions on a hypercubic lattice.

The aim is to generate independent configurations distributed as

$$P(U_l)=\exp[-\beta S(U_l)], \tag{2}$$

with

$$S(U_l)=\sum_{P\supset U_l}\left[1-\tfrac{1}{2}\operatorname{tr}(U_P)\right]. \tag{3}$$

Here, the U_l's are link variables and U_P is the ordered product of the link variables around a unit lattice square (plaquette). It was shown in Ref. 2 that an overrelaxation parameter can be introduced into the Markov evolution by the generation of new values U_l' from U_l according to the probability function

$$W(U_l\rightarrow U_l')=\mathcal{N}\exp\left\{\frac{1-\omega}{\omega(2-\omega)}\sum_{P\supset U_l}\beta[1-\tfrac{1}{2}\operatorname{tr}(U_PU_P')]-\frac{1}{\omega}\sum_{P\supset U_l}\beta[1-\tfrac{1}{2}\operatorname{tr}(U_P')]-\frac{1-\omega}{\omega}\sum_{P\supset U_l}\beta[1-\tfrac{1}{2}\operatorname{tr}(U_P)]\right\}, \tag{4}$$

with $\mathcal{N}$ fixed by the normalization condition

$$\int d[U_l']\,W(U_l\rightarrow U_l')=1. \tag{5}$$

The replacement $U_l\rightarrow U_l'$ is accepted with probability

$$P_A=\min\left\{1,\frac{W(U_l'\rightarrow U_l)\exp[-\beta S(U_l')]}{W(U_l\rightarrow U_l')\exp[-\beta S(U_l)]}\right\}. \tag{6}$$

It is easy to see that this procedure satisfies detailed balance with respect to the desired Boltzmann distribution of Eq. (2).[3]

P_A can be simplified to

$$P_A=\min\left\{1,\frac{\mathcal{N}'}{\mathcal{N}}\right\}, \tag{7}$$

and $\mathcal{N}$ can be explicitly evaluated [for the SU(2) case] to be

$$\mathcal{N}=Ae^{-R}\frac{k}{I_1(2k)}, \tag{8}$$

with A a numerical constant, I_1 the modified Bessel function of first order,

$$k=(\operatorname{Det}M)^{1/2}, \tag{9}$$

$$M=\frac{\beta}{2\omega}\sum_{P\supset U_l}U_S\left[-1+\frac{1-\omega}{2-\omega}U_P\right], \tag{10}$$

VOLUME 62, NUMBER 2 PHYSICAL REVIEW LETTERS 9 JANUARY 1989

where $U_P = U_l U_S$, and

$$R = -\frac{1}{\omega(2-\omega)} \sum_{P\supset U_l} \beta - \frac{1-\omega}{\omega} \sum_{P\supset U_l} \beta[1-\tfrac{1}{2}\operatorname{tr}(U_P)]\,. \tag{11}$$

The simulation procedure is to preselect a U_l' with Eq. (4) and then to use Eq. (7) to decide whether or not to make the replacement $U_l \to U_l'$. Since W of Eq. (4) is linear in U_l' one can generate it by a heat-bath algorithm (we use the method of Kennedy and Pendelton[4]). Also, the whole procedure is readily vectorizable. Finally, although the formulas above [Eqs. (8)-(11)] look formidable, they are easy to implement in a real simulation since R and M are readily available in the course of a standard heat-bath update. The increase in computer time for each update over the $\omega=1$ case was about 20% to 25% in our study. The main extra work is in the generation of I_1, which can be vectorized.

Suppose one is measuring the expectation value of some operator $\mathcal{O}$. Consider first the case when $\mathcal{O}$ is the plaquette operator $\frac{1}{2}\operatorname{tr}(U_P)$. This operator probes the field on scales of the order of a unit lattice spacing. When $\xi \ll 1$ (i.e., β is small), the decorrelation time will be small and no improvement will be necessary since one update is sufficient to decorrelate on length scales of order unity. On the other hand, when $\xi \gg 1$, the plaquette is a very local operator and again should decorrelate quickly. The worst case for the plaquette is when $\xi \sim 1$. Generalizing from this, if an operator gets dominant contributions from variations of the field over length scales of the order of l, then its decorrelation time will be maximum when $\xi \sim l$. What this argument suggests is that the optimum value of the overrelaxation parameter ω may have to be determined separately for each operator one wants to measure. This is very different from the differential-equation case where there is a unique meaning to the "best value of ω," based on optimization of the decorrelation of the longest-wavelength modes.[5] Indeed, there may even be operators that are sensitive to many length scales whose decorrelation time *cannot* be improved by overrelaxation.

Let us measure the expectation value $E(I,\beta) = \langle \mathcal{O} \rangle$ of $\mathcal{O}$ from N successive configurations after I_0 thermalizing updates of the lattice. I labels the iteration number, $I = I_0+1, I_0+2, \ldots, I_0+N$. One defines the autocorrelation function $C(J,\beta)$ as

$$C(J,\beta) = \frac{D(J,\beta)}{D(0,\beta)}\,, \tag{12a}$$

$$D(J,\beta) = \sum_{I=I_0+1}^{I_0+N-J} [E(I,\beta) - \bar{E}(\beta)] \times [E(I+J,\beta) - \bar{E}(\beta)]\,. \tag{12b}$$

Here $\bar{E}(\beta) = N^{-1} \sum_{I=I_0+1}^{I_0+N} E(I,\beta)$ is the average value of E in the simulation.

If the system is dominated by a single decorrelation time τ, then for large J,

$$C(J,\beta) \simeq e^{-J/\tau(\beta)}\,. \tag{13}$$

In Fig. 1 we show $C(5,\beta)$ and $C(10,\beta)$ for the plaquette as a function of β at $\omega=1$ on 4^4, 6^4, and 8^4 lattices. Note that ξ is a monotonically increasing function of β, and that for fixed J, τ is a monotonically increasing function of C [see Eq. (13)]. Hence, Fig. 1 is a practical realization of the argument presented above which predicted a peak in τ as a function of ξ. Indeed, it is known that in the SU(2) theory, $\xi \sim 1$ near $\beta = 2.2$ and this is just where the peak is. Incidentally, this peak is also correlated with the peak in the specific heat and the location of the crossover from strong to weak coupling.

Figure 1 also shows the analogous situation for the magnitude of the Polyakov loop. Note that now, $C(5,\beta)$ and $C(10,\beta)$ become nonzero only for larger β values and the "turn on" value of β is different for different lattice sizes. Note also that there is no peak in $C(J,\beta)$ for the Polyakov loop as a function of β. This implies that the Polyakov loop gets contributions from all length scales on the lattice up to the lattice size. One might suspect that such operators might not have their correlation time decreased by overrelaxation methods, because as shown in Refs. 2 and 5, optimized overrelaxation speeds up the decorrelation of the longest-wavelength modes while simultaneously slowing down the decorrelation of the shortest-wavelength modes.

To study overrelaxation, we first studied $C(5,\beta)$ and $C(10,\beta)$ for the plaquette at $\beta = 2.2$ on lattices of size 4^4, 6^4, 8^4, and 12^4. The overrelaxation was done by one

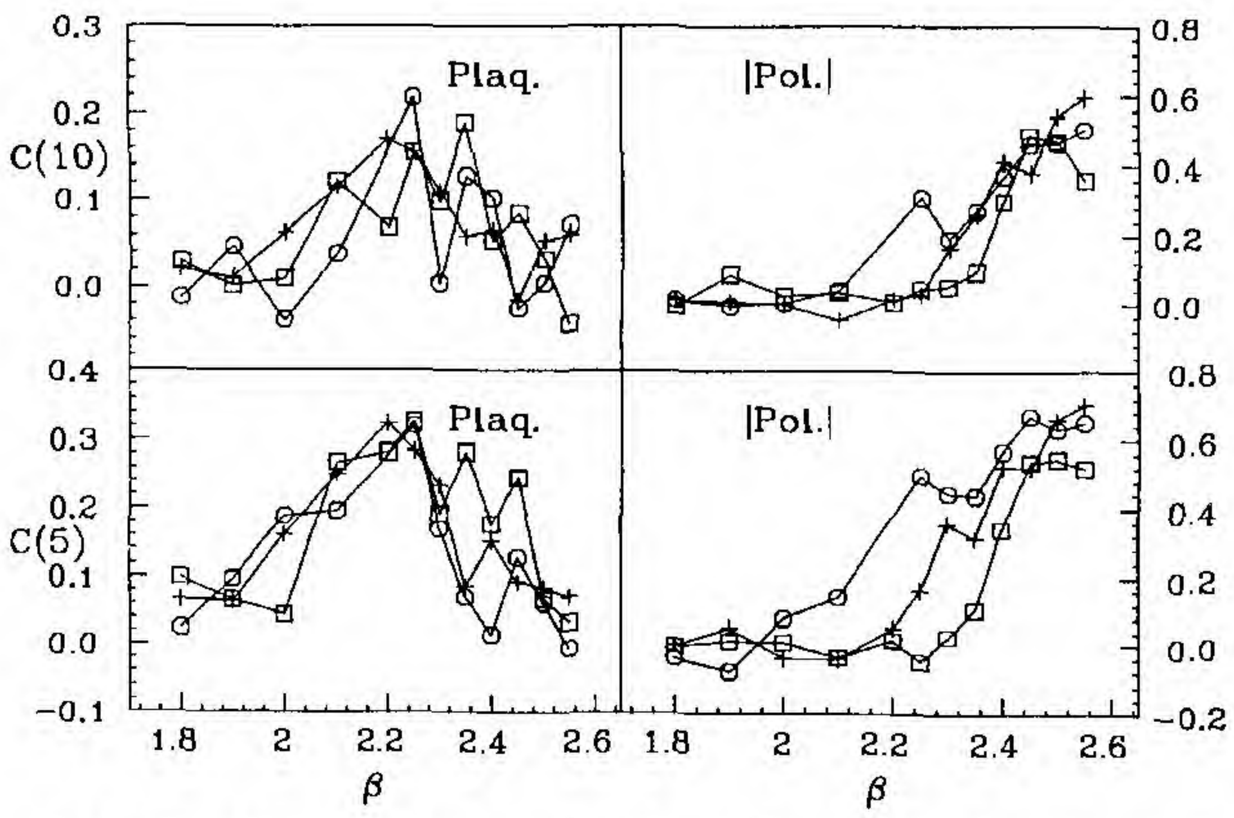


FIG. 1. Autocorrelation function $C(J)$ at $J=5$ and $J=10$ for the plaquette (Plaq.) and the magnitude of the Polyakov loop (|Pol.|) at $\omega=1$ as a function of β. Circles, plusses, and squares represent data for 4^4, 6^4, and 8^4 lattices, respectively. Note the peak in C near $\beta=2.2$ for the plaquette and its absence for the Polyakov loop.

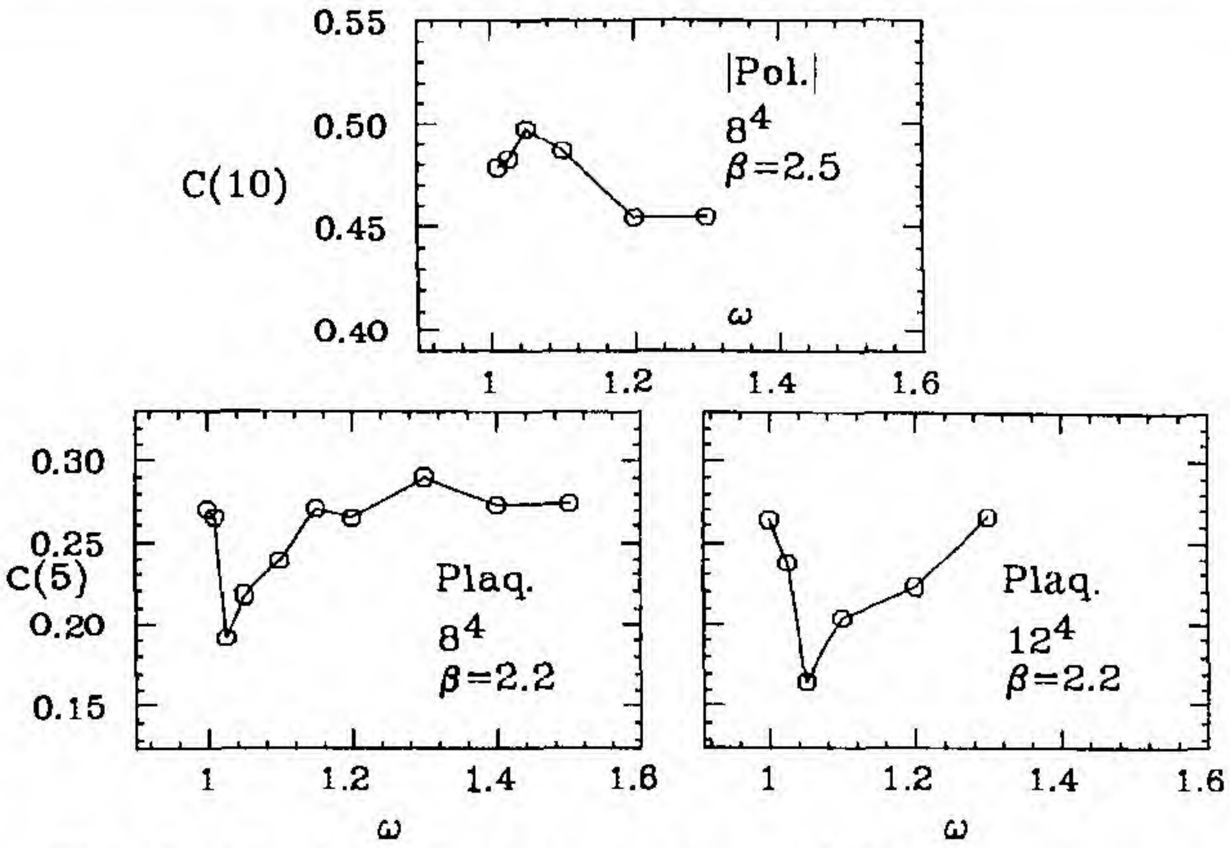


FIG. 2. The bottom two plots show $C(5)$ for the plaquette on 8^4 and 12^4 lattices at $\beta=2.2$ as a function of ω. Note the sharp improvement in decorrelation time near $\omega=1.025$ for an 8^4 lattice and near $\omega=1.05$ for a 12^4 lattice. The upper plot shows $C(10)$ for the Polyakov loop. There is no improvement in decorrelation time in this case.

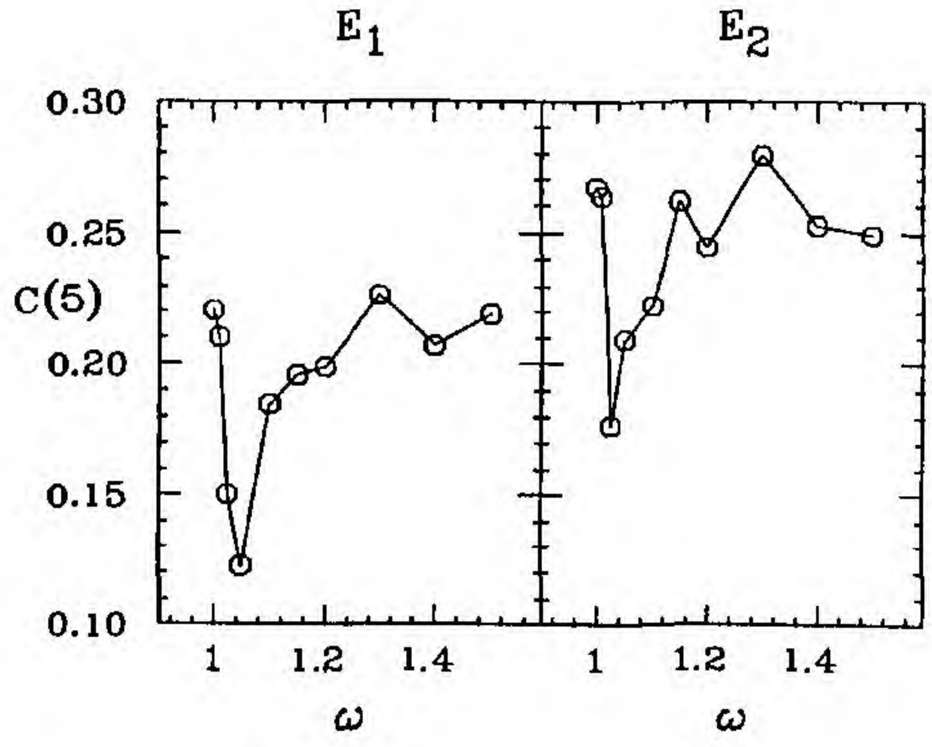


FIG. 3. $C(5)$ for the order parameters E_1 and E_2 [see Eq. (14) of the text] at $\beta=2.2$ on an 8^4 lattice. Note the improvement near $\omega=1.025$ in all cases.

sweep of the lattice with $\omega=1$ followed by four sweeps during which the overrelaxation parameter was increased linearly to its target value. 1000 such sequences of five sweeps were done at each β value starting from an ordered configuration ($U_l=1\ \forall\, l$) and $C(5,\beta)$ and $C(10,\beta)$ were computed from the last 4000 lattices generated. The idea of changing ω in this sawtooth fashion was motivated by the Chebyshev acceleration schemes in partial-differential-equation theory.[6] The results are shown in Fig. 2. Only results for an 8^4 and 12^4 lattice are shown because we did not see any improvement for smaller lattice sizes. The optimum value of ω is cleary near $\omega=1.025$ for an 8^4 lattice and near $\omega=1.05$ for a 12^4 lattice. If τ is estimated from Eq. (13), the improvement is a factor of 2 or 3 in decorrelation time for the average plaquette. Figure 2 also shows results for a similar simulation for the magnitude of the Polyakov loop on an 8^4 lattice at $\beta=2.5$. Here, there does not seem to be an improvement.

We have also computed, at $\beta=2.2$, the autocorrelation function for the following two order parameters: The average adjoint plaquette,

$$E_1(\beta)=\tfrac{1}{4}\langle[\mathrm{tr}(U_P)]^2\rangle, \tag{14a}$$

and the plaquette-plaquette correlation function at distance unity,

$$E_2(\beta)=\tfrac{1}{4}\langle \mathrm{tr}[U_P(x)]\mathrm{tr}[U_P(x+\hat\mu)]\rangle. \tag{14b}$$

Here $\hat\mu$ denotes a lattice axis direction and the correlation of Eq. (14b) is computed for face-to-face plaquettes. Since these are both order parameters dominated by short-distance effects, they should improve if $E_0=\frac{1}{2}\times\langle\mathrm{tr}(U_P)\rangle$. This is indeed true and is shown in Fig. 3.

Finally, one must address the issue of acceptance rates. Since the optimum value of ω is close to unity, the acceptance rate at the optimum ω is close to that at $\omega=1$. In Fig. 4, we show the acceptance rate relative to $\omega=1$ as a function of β at $\omega=1.4$ and as a function of ω at $\beta=2.5$ on an 8^4 lattice. We found that the acceptance rate was almost independent of lattice size.

In the classical[6] (or free-field[2]) analysis, the optimum ω is related to the lattice size for large lattices by

$$\omega_{\rm opt}\approx\frac{2}{1+C/L}, \tag{15}$$

with C an appropriate constant. Thus, as the lattice size L increases the optimum ω also increases, approaching $\omega_{\rm opt}=2$ in the large L-limit. This trend can be seen in the two lattices for which we have determined $\omega_{\rm opt}$ for the plaquette, but Eq. (15) only roughly holds: The calculation of C from $C=L(2/\omega_{\rm opt}-1)$ gives $C\approx 8,11$ for the $L=8,12$ simulations. Since C/L is of order unity for

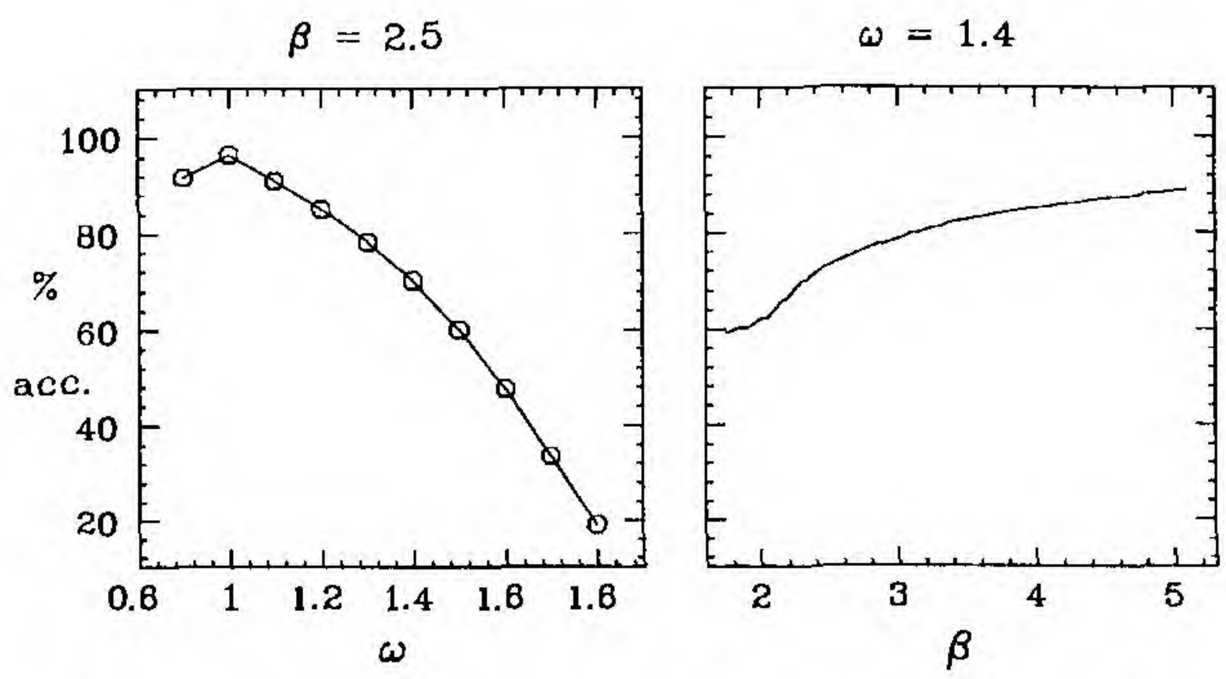


FIG. 4. The left-hand plot shows the acceptance at $\beta=2.5$ as a function of ω on an 8^4 lattice. The right-hand plot shows the acceptance at $\omega=1.4$ (relative to the acceptance at $\omega=1$) as a function of β, also on an 8^4 lattice.

VOLUME 62, NUMBER 2 PHYSICAL REVIEW LETTERS 9 JANUARY 1989

these lattices, the $1/L^2$ corrections to Eq. (15) can be expected to be important, and so the variation in the fitted C values is perhaps not surprising. The value of C obtained is considerably larger than the classical value $C \approx 3$ for similar boundary conditions, and in accordance with our discussion above, is likely to vary with the order parameter being probed.

In summary, we have presented a study of overrelaxation which shows that it can be a useful tool in the study of certain order parameters. Left unanswered are the questions of how to determine *a priori* which measurements benefit from overrelaxation and which do not, and how the systematics of optimizing ω differs from the classical case. It will be important to resolve these issues, by further empirical study or theoretical analysis,[5] before overrelaxation methods are applied to large-scale simulations in gauge theory.

The work of S. A. and G. B. was supported in part by the U.S. Departmetn of Energy under Grant No. DE-AC02-76ERO2220 and Contract No. DE-FC05-85ER250000, respectively. We would like to thank Herbert Neuberger for many illuminating conversations. G. B. thanks the Institute for Advanced Study for hospitality during the inception and completion of this work and Khalil Bitar for discussions.

[1]S. L. Adler, Phys. Rev. D **23**, 2901 (1981); C. Whitmer, Phys. Rev. D **29**, 306 (1984), and thesis, Princeton University (unpublished); M. Creutz, Phys. Rev. D **36**, 515 (1987); F. R. Brown and T. J. Woch, Phys. Rev. Lett. **58**, 2394 (1987).

[2]S. L. Adler, Phys. Rev. D **37**, 458 (1988), and **38**, 1349 (1988).

[3]One might imagine using the W of Eq. (4) as a Markov transition probability. However, the resulting equilibrium distribution is then not the desired one $P = \exp[-\beta S(U_l)]$, but rather one with action S' differing from S by terms of order a^2, with a the lattice spacing. For small a one might argue formally that this would be satisfactory. However, by direct simulation, one finds that S' has spurious first-order transitions at finite β for $\omega > 1$ which prevent taking the continuum ($\beta \to \infty$) limit.

[4]A. D. Kennedy and B. J. Pendelton, Phys. Lett. **156B**, 393 (1985).

[5]H. Neuberger, Phys. Lett. B **207**, 461 (1988).

[6]For a detailed discussion of the choice of ω in the differential-equation case, see R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981), pp. 174–181; R. S. Varga, *Matrix Iterative Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1962), Chap. 5.

Nuclear Physics B (Proc. Suppl.) 9 (1989) 437-446
North-Holland, Amsterdam

ALGORITHMS FOR PURE GAUGE THEORY

Stephen L. ADLER

The Institute for Advanced Study, Princeton, New Jersey 08540, USA

In this talk I review recent progress on algorithms for pure gauge theory, concentrating mainly on overrelaxation methods, with some brief comments on Fourier accelerated Langevin and multigrid algorithms at the end.

1. THE CRITICAL SLOWING DOWN PROBLEM

Let m be the mass gap or characteristic mass scale of the physics being studied, and let $a = L^{-1}$ be the lattice spacing. The ratio of the characteristic length m^{-1} to a is called the correlation length ξ,

$$\xi = m^{-1}/a = L/m \ , \tag{1}$$

and becomes infinite in the continuum limit $L \to \infty$. In a Monte Carlo calculation, noise introduced at x, t must diffuse to x', $|x' - x| \sim \xi$ to yield an independent lattice configuration; this requires a characteristic "decorrelation time" or "autocorrelation time" τ. In the free field case $\tau \sim \xi^2$; more generally

$$\tau \sim \xi^z \ , \tag{2}$$

with z a dynamical critical exponent. Let W be the computational work needed to get an independent configuration. For an ideal "fast" algorithm

$$W \sim L^d \times \text{logs of } L \ , \tag{3a}$$

but from eq. (2) we in fact have

$$\frac{W}{L^d} \sim \xi^z \sim L^z \text{ as } L \to \infty \ , \tag{3b}$$

and so if $z > 0$, W is much larger than ideal for large lattices. This is the "critical slowing down problem;" ways of alleviating it are the subject matter of this talk.

2. RELATION BETWEEN THERMALIZATION AND MINIMIZATION

In Monte Carlo we study the partition function

$$z = \int d[U] \, e^{-\beta S(U)} \ , \tag{4}$$

and corresponding Green's functions, by generating a sequence of configurations $\{U_i\}$ distributed according to the probability distribution

$$P(U) = e^{-\beta S(U)} \ . \tag{5}$$

Regarding β as T^{-1} with T a "temperature," let us consider the $T \to 0$ or $\beta \to \infty$ limit of eq. (4). In this limit fluctuations in U are frozen out, and z is dominated by the minimum of $S(U)$ at which

$$\delta_U \, S(U) = 0 \ . \tag{6}$$

Any Monte Carlo algorithm, in the $\beta \to \infty$ limit, becomes an algorithm for the minimization problem of eq. (6); conversely, we may expect that certain algorithms for eq. (6) can be generalized to finite β to provide useful new Monte Carlo algorithms.

3. THE GAUSSIAN OR FREE FIELD MODEL

To develop the relation between thermalization and minimization, it is instructive to consider in detail the Gaussian or free-field model.

3.1. Gaussian minimization

Let $\tilde{L} = \tilde{L}^T$ be a real, symmetric $\ell \times \ell$ matrix and u, f real ℓ-dimensional vectors, and consider the action

$$S(u) = \frac{1}{2} u^T \tilde{L} u - f^T u \ , \tag{7}$$

for which the variational equation $\delta_u S(u) = 0$ gives the linear system

$$\tilde{L} u = f \ . \tag{8}$$

A general linear iteration for solving eq. (8) is defined by

$$u_{n+1} = M u_n + N f \ , \tag{9}$$

with M, N a "splitting" of $\tilde{L}$ which obeys

$$1 = M + N\tilde{L} . \quad (10)$$

The condition of eq. (10) guarantees that if $u_{n+1} \to u_n \to u_\infty$, then

$$u_\infty = Mu_\infty + Nf \quad (11)$$

is equivalent to $\tilde{L}u_\infty = f$, so that u_∞ is the desired solution of eq. (8). Subtracting eq. (11) from eq. (9), we have

$$(u_{n+1} - u_\infty) = M(u_n - u_\infty) ; \quad (12)$$

hence errors in the initial guess decay as M^n, and the dominant error decays as ρ^n, where

$$\rho = \text{"spectral radius"} = \sup_i |M_i| ,$$
$$M_i = \text{eigenvalue of } M . \quad (13)$$

The iteration converges if $\rho < 1$ which requires $M_i = 1 - \varepsilon_i$, $2 > \varepsilon_i > 0$ and the dominant error clearly decays with characteristic time

$$\tau_{error\ decay} \sim \varepsilon_{min}^{-1} . \quad (14)$$

But from eq. (10), $1 - M = N\tilde{L}$ and hence we expect

$$\varepsilon_{min} \sim \tilde{L}_{min} = \text{smallest eigenvalue of } \tilde{L} , \quad (15)$$

i.e., the smallest eigenvalue of $\tilde{L}$ gives the slowest decaying errors. For example, in the dimension d free field case, where $\tilde{L}$ is the discretization of ∇_d^2, we have $\tilde{L}_{min} \sim a^2 \sim L^{-2}$, which implies $\tau_{error\ decay} \sim L^2$, or $z = 2$.

3.2. Gaussian thermalization

The minimization algorithm of eq. (9) for the action of eq. (7) can be generalized into a Monte Carlo algorithm as follows.[1] Let $P(u \to u')$ be the normalized probability for the transition from $u = u_n$ to $u' = u_{n+1}$, given by

$$P(u \to u') = \left(\tfrac{\beta}{\pi}\right)^{\ell/2} (\det \Gamma)^{1/2} \times$$
$$\exp\left[-(u' - Mu - Nf)^T \beta\, \Gamma(u' - Mu - Nf)\right] , \quad (16a)$$

with $\Gamma = \Gamma^T$ a generalized inverse temperature matrix related to $\tilde{L}$ and M by

$$\Gamma = \frac{1}{2}\left(\tilde{L}^{-1} - M\tilde{L}^{-1} M^T\right)^{-1} . \quad (16b)$$

A straightforward calculation shows that $P(u \to u')$ satisfies the homogeneous state condition

$$\int du\, e^{-\beta S(u)}\, P(u \to u') = e^{-\beta S(u')} , \quad (17)$$

which is necessary for Monte Carlo; P satisfies the stronger detailed balance condition

$$e^{-\beta S(u)}\, P(u \to u') = e^{-\beta S(u')}\, P(u' \to u) , \quad (18)$$

which is sufficient but not necessary for Monte Carlo, if

$$M^T\tilde{L} = \tilde{L}M \iff N = N^T . \quad (19)$$

3.3. Stochastic difference equation form of Gaussian thermalization

Let O be the real, orthogonal matrix which diagonalizes Γ, so that $O^T\, \Gamma O = \gamma =$ diagonal . Then eq. (16) can be written as a stochastic difference equation

$$u' = Mu + Nf + \frac{1}{2\beta^{1/2}}\, O\, \gamma^{-1/2}\eta , \quad (20)$$

with η_i Gaussian noise normalized as $\langle\eta_i\eta_j\rangle_\eta = 2\delta_{ij}$. Iterating eq. (20) n times we see that the autocorrelation $\langle u_{N+n}u_N\rangle$, averaged over u_N, behaves as

$$\langle u_{N+n}u_N\rangle_{\eta,u_N} \sim M^n . \quad (21)$$

Hence, as first shown by Goodman and Sokal,[2] in the Gaussian case the stochastic decorrelation time τ is the same as the deterministic error decay time $\tau_{error\ decay}$, and thus from eq. (14) we have

$$\tau \sim \varepsilon_{min}^{-1} . \quad (22)$$

3.4. Local algorithms—minimization

Consider now the specialization of the iteration of eq. (9) which acts only on a single node variable u_k. As a function of this node,

$$S(u) = A_k(u_k - C_k)^2 + B_k ,$$
$$A_k,\ B_k,\ C_k \text{ independent of } u_k , \quad (23)$$

and for stability we assume $A_k > 0$. The simplest local minimization algorithm is the Gauss-Seidel iteration $u_k \to u_k' = C_k$, in which one replaces the old value of u_k with the value C_k which minimizes the action. The Gauss-Seidel algorithm discards the information contained in the old value of u_k; a more efficient algorithm, when coherent effects over the whole lattice are taken into account, is to replace u_k by a linear combination of C_k and the old value,

$$u_k \to u_k' = \omega\, C_k + (1 - \omega)u_k . \quad (24)$$

Equation (24) is called the overrelaxed Gauss-Seidel algorithm, and ω is called the overrelaxation parameter. The iteration of eq. (24) will cause the action to decrease if

$$0 \leq S(u_k, \ldots) - S(u'_k, \ldots) = A_k(u_k - u'_k)^2 \left(\tfrac{2}{\omega} - 1\right) , \tag{25}$$

which requires $0 < \omega < 2$. Theory shows that convergence is optimum for a value ω_{opt} which scales, for large L, as

$$\omega_{opt} \approx \frac{2}{1 + \frac{C_{opt}}{L}} , \tag{26}$$

with C_{opt} a constant.

3.5. Local algorithms—thermalization

Since the algorithms of sect. 3.4 are special cases of eq. (9), they will have thermal generalizations. The thermal version of the Gauss-Seidel algorithm is the heat bath, for which $P(u_k \to u'_k)$ is

$$\begin{aligned} P(u_k \to u'_k) &= \left(\tfrac{\beta A_k}{\pi}\right)^{1/2} e^{-\beta A_k (u'_k - C_k)^2} \\ &= \text{const} \times e^{-\beta S(u'_k, \ldots)} , \end{aligned} \tag{27}$$

and is independent of the old node value u_k. The thermal generalization of the overrelaxed Gauss-Seidel algorithm, proposed by Adler[3] in 1981, is

$$\begin{aligned} P(u_k \to u'_k) &= \left[\tfrac{\beta A_k}{\pi\omega(2-\omega)}\right]^{1/2} \times \\ &\exp\left\{-\tfrac{\beta A_k}{\omega(2-\omega)} \left[u'_k - \omega\, C_k - (1-\omega) u_k\right]^2\right\} , \end{aligned} \tag{28}$$

which satisfies detailed balance with respect to the action S of eq. (23). Again, ω must lie in the range $0 < \omega < 2$. For $1 < \omega < 2$, the mean u'_k in the distribution of eq. (28) is $\bar{u}'_k = C_k + (\omega - 1)(C_k - u_k)$ and is displaced beyond C_k in the opposite direction from the old value u_k, hence the name overrelaxation. In addition to displacing the mean, eq. (28) reduces the width by a factor of $\sqrt{\omega(2-\omega)} < 1$ as compared with heat bath. The case $\omega = 2$ is of special interest; it corresponds to a deterministic reflection of u_k around C_k, since for $\omega = 2$ the width of the distribution is zero and the mean is $\bar{u}'_k = C_k + C_k - u_k$. The corresponding action change [eq. (25)] is $\Delta S = 0$ for $\omega = 2$, and so the $\omega = 2$ limit of eq. (28) is microcanonical, or action conserving. This fact is exploited, as discussed below, in the Brown and Woch, Creutz $SU(n)$ implementation of overrelaxation.

3.6. Eigenvalue analysis of the iteration matrix M for ∇^2_d

In the free field case, the eigenvalue spectrum of M for the overrelaxed algorithm has been calculated.[4,5,6] It is convenient to Fourier analyze and look at the submatrix M_k for wave number k; in Neuberger's[5] calculation, using red-black ordering with periodic boundary conditions, M_k is a 2×2 matrix and the analysis is particularly simple. For the Gauss-Seidel or heat bath algorithms one finds

$$\tau_k^{-1} = \epsilon_k = \text{const} \times a^2 k^2 , \tag{29}$$

where

$$k_{min} = O(1),\ k_{max} = O(L) , \tag{30a}$$

and so

$$\begin{aligned} \tau_{\text{long wavelength}} &\sim L^2 , \\ \tau_{\text{short wavelength}} &\sim 1 , \end{aligned} \tag{30b}$$

and the critical exponent z defined by eqs. (1) and (2) is $z = 2$. For the overrelaxed Gauss-Seidel and heat bath algorithms, there is an optimum value $\omega = \omega_{opt}$ for which

$$\tau_k^{-1} = \epsilon_k = \text{const} \times k_{min} , \quad \text{all } k , \tag{31a}$$

and so

$$\tau_{\text{all wavelengths}} \sim L \tag{31b}$$

which implies $z = 1$! Note that while overrelaxation speeds up the decorrelation of the long wavelength modes, it slows down the decorrelation of the shortest modes. As noted above, for large L the optimum ω scales as $\omega_{opt} = 2/(1+C_{opt}/L)$, and the detailed analysis shows $C_{opt} \sim k_{min}/d^{1/2}$. More generally, for the family scaling towards 2 for large L as $\omega = 2/(1 + C/L)$, $C \neq C_{opt}$, one still has $z = 1$, but the coefficient of proportionality τ/ξ is not optimal. Thus, in the free-field case, overrelaxation reduces $z = 2$ to $z = 1$, and goes half-way towards solving the critical slowing down problem.

3.7. ω-fixing

In the minimization problem, it can be shown[7] that $\omega = 1$ maximizes the initial rate of error decay, while the choice $\omega = \omega_{opt}$ maximizes the asymptotic rate of error decay for large computational time, but can actually increase errors for some finite number of early iterations. Hence the optimal strategy to get the smallest error at each stage of computation is to use for sweep i an ω value ω_i, with $\omega_0 = 1$, $1 < \omega_i < \omega_{opt}$ for $i > 1$, and $\omega_{i \to \infty} = \omega_{opt}$; under certain assumptions about the errors, the sequence ω_i can be explicitly constructed in terms of Chebyshev polynomials. In the Monte Carlo case, the noise injected at each update behaves as a new initial error term, and so the decorrelation strategy suggested by Chebyshev overre-

laxation is to use a "sawtoothed" sequence of ω's, $1,\ \omega_1,\ldots,\omega_N,\ 1,\ \omega_1,\ldots\omega_N,\ 1,\ \omega_1,\ldots,\omega_N,\ldots$, with $\omega_i, 1 \le i \le N$ increasing towards ω_{opt}, and with the sequence length N chosen of order the effective decorrelation time τ. (Note that this gives an implicit, self-consistent specification of N.) Then taking one lattice configuration for measurement per sawtooth sweep should give the maximum number of statistically independent configurations per unit of computational work. The "OR n" methods discussed below will be seen to be analogs of the sawtooth scheme.

3.8. Multiquadratic generalizations

The algorithm of eq. (28) is valid for any action which, as a function of each individual node variable u_k, has the form of eq. (23). This defines the class of *multiquadratic* actions, which is larger than the class of Gaussian actions. Some interesting interacting multiquadratic actions are:

- The continuum Yang-Mills action $S = \frac{1}{4}(F^a_{\mu\nu})^2$, which is multiquadratic[8] because the interaction term in $F^a_{\mu\nu}$ is an outer product $f^{abc}A^b_\mu A^c_\nu$;
- Non-compact discrete Yang-Mills, for the same reason;
- The discretized ϕ^4 and Higgs actions with ϕ^4 point split[9] according to $\phi^4(x) \to \phi^2(x)\phi^2(x+a\hat{\mu})$, with $\hat{\mu}$ a unit lattice displacement.

4. NON-MULTIQUADRATIC IMPLEMENTATIONS OF OVERRELAXATION

Most non-multiquadratic implementations of overrelaxation make use of the generalized Metropolis algorithm, consisting of the following two steps:

1. Given u, pick a trial u' with normalized probability distribution $W(u \to u')$.
2. Accept u' with conditional probability

$$P_A = \min\left\{1,\ \frac{W(u' \to u)}{W(u \to u')}\ \frac{e^{-\beta S(u')}}{e^{-\beta S(u)}}\right\} . \qquad (32)$$

This procedure gives an overall transition probability $P(u \to u')$ which satisfies the detailed balance condition of eq. (18). I will group the algorithms to be discussed into two basic types, (a) microcanonical analog algorithms, and (b) tunable-ω algorithms.

4.1. Microcanonical analog algorithms

The simplest implementation of overrelaxation for $SU(n)$ gauge theories is the microcanonical analog algorithm proposed by Brown and Woch[10] and Creutz.[11] Let U_ℓ be the link being updated, and let S_W be the sum of terms in the Wilson action which depend on U_ℓ:

$$\beta S_W(U_\ell) = \text{const} - \frac{\beta}{n}\ \text{Re Tr}\left(U_\ell \sum_{\text{staples}} U_S\right) . \qquad (33)$$

Let U_0 be the value of U_ℓ which minimizes S_W; in terms of U_0 and the old value of the link U_ℓ, we wish to construct an analog of the $\omega = 2$ limit of the multiquadratic algorithm of eq. (28), which we rewrite as

$$u'_k = C_k - u_k + C_k . \qquad (34a)$$

Exponentiating eq. (34a) gives

$$e^{i\,u'_k} = e^{i\,C_k}\left(e^{i\,u_k}\right)^{-1} e^{i\,C_k} , \qquad (34b)$$

which suggests the unitary group analog

$$U'_\ell = U_0\, U_\ell^{-1}\, U_0 . \qquad (35)$$

The algorithm of Brown and Woch and Creutz consists of using eq. (35) as a Metropolis trial selection; since the inversion of eq. (35),

$$U_\ell = U_0\,(U'_\ell)^{-1}\, U_0 , \qquad (36)$$

is just eq. (35) with U_ℓ and U'_ℓ interchanged, the ratio $W(U'_\ell \to U_\ell)/W(U_\ell \to U'_\ell)$ in eq. (32) reduces to unity, and detailed balance with the Wilson action of eq. (33) is satisfied by accepting the trial selection of eq. (35) with the conditional probability

$$P_A = \min\left\{1,\ e^{-\beta[S_W(U'_\ell) - S_W(U_\ell)]}\right\} . \qquad (37)$$

For $SU(n)$ with $n \ge 3$, the acceptance probability P_A is smaller than unity. The case of $SU(2)$ is special, since the sum over staples in eq. (33) is then proportional to an $SU(2)$ group element which we call U_0^{-1},

$$\sum_{\text{staples}} U_S = k\, U_0^{-1} ,$$

$$\beta S_W(U_\ell) = \text{const} - \frac{\beta k}{n}\ \text{Tr}(U_\ell U_0^{-1}) . \qquad (38)$$

We see from eq. (38) that indeed $U_\ell = U_0$ minimizes $S_W(U_\ell)$, and since

$$\begin{aligned}\mathrm{Tr}(U_\ell' U_0^{-1}) &= \mathrm{Tr}(U_0 U_\ell^{-1} U_0) U_0^{-1} = \mathrm{Tr}(U_0 U_\ell^{-1}) \\ &= \mathrm{Tr}(U_0 U_\ell^{-1})^{-1} = \mathrm{Tr}\ U_\ell U_0^{-1} \ ,\end{aligned} \tag{39}$$

we have $S_W(U_\ell') = S_W(U_\ell)$. So for $SU(2)$ we have $P_A = 1$, and the Brown and Woch, Creutz algorithm is microcanonical. To get ergodicity one adds standard Metropolis steps, defining "OR n Metropolis" as follows:

OR 0 Metropolis =

Brown and Woch, Creutz algorithm

of eqs. (35) and (37); (40)

OR n Metropolis =

OR 0 $+n$ standard Metropolis steps.

Clearly, OR n is similar in spirit to the "sawtoothing" procedure discussed in sect. 3.7.

In analogy with the above $SU(n)$ algorithm, Gupta, et al.[12] have given an overrelaxed Metropolis algorithm for the XY model. The action here is

$$\begin{aligned} S &= \mathrm{const} - \beta \textstyle\sum_{\langle ij \rangle} \cos(\theta_i - \theta_j) \\ &= \mathrm{const} - \beta\, k \textstyle\sum_i \mathrm{Re}(S_i \Sigma_i^\dagger) \ , \end{aligned} \tag{41}$$

$$S_i = e^{i\theta_i},\ k\, \Sigma_i = \textstyle\sum_{j \in \langle ij \rangle} S_j \ ,$$

where $\langle ij \rangle$ indicates the restriction of the sum to nearest neighbor pairs. eq. (41) has the same structure as eqs. (33) and (38), so a corresponding overrelaxed, microcanonical algorithm is

$$S_i \to S_i' = \Sigma_i\, S_i^{-1}\, \Sigma_i \ . \tag{42}$$

4.2. Tunable-ω algorithms

Adler[4,13] has given an overrelaxed algorithm for $SU(n)$ incorporating a tunable ω parameter. The algorithm is constructed by first doing a non-compact discretization of $SU(n)$ and using eq. (28) to write the corresponding overrelaxed transition probability $W(u \to u')$. One then writes the Wilson lattice gauge theory transcription of W, which takes the following form. Let ℓ be the link being updated, with U_ℓ the old value, U_ℓ' the new value, and U_S the staple joining with U_ℓ in the plaquette P. In terms of $U_P = U_\ell U_S$, $U_P' = U_\ell' U_S$ we have

$$\begin{aligned} W[U_\ell \to U_\ell'] = \mathcal{N} \times \\ \exp\Big\{ \tfrac{1-\omega}{\omega(2-\omega)} \textstyle\sum_{P \supset \ell} \beta \left[1 - \tfrac{1}{n}\, \mathrm{Re\ Tr}\ (U_P U_P')\right] \\ - \tfrac{1}{\omega} \textstyle\sum_{P \supset \ell}\ \beta \left[1 - \tfrac{1}{n}\, \mathrm{Re\ Tr}\ U_P'\right] \\ - \tfrac{1-\omega}{\omega} \textstyle\sum_{P \supset \ell}\ \beta \left[1 - \tfrac{1}{n}\, \mathrm{Re\ Tr}\ U_P\right] \Big\} \ , \end{aligned} \tag{43a}$$

with the normalization constant $\mathcal{N}$ fixed by

$$\int d[U_\ell']\ W[U_\ell \to U_\ell'] = 1 \ . \tag{43b}$$

We then use W as a Metropolis preselection, and accept U_ℓ' with the conditional probability

$$\begin{aligned} P_A &= \min\left\{1,\ \tfrac{W[U_\ell' \to U_\ell]}{W[U_\ell \to U_\ell']}\ e^{-\beta[S_W(U_\ell') - S_W(U_\ell)]}\right\} \\ &= \min\{1,\ \mathcal{N}'/\mathcal{N}\} = \min\{1,\ 1 + O(a^2)\} \ , \end{aligned} \tag{44}$$

giving an efficient algorithm for which $P_A \to 1$ in the continuum limit. In the case of $SU(2)$, the normalization $\mathcal{N}$ can be readily calculated in closed form.[13]

An alternative tunable-ω algorithm, for the special case of $SU(2)$, has been given by Brown and Woch.[10] They proceed by mapping $SU(2)$ onto R^3 in such a way that the Wilson action transforms into a unit Gaussian. They then do an overrelaxed heat bath update using eq. (28), and finally remap back to $SU(2)$. Explicitly, they start from eq. (38) and set

$$U_\ell = e^{i\, \theta\, \hat{r} \cdot \vec{\sigma}} U_0 \ , \tag{45}$$

with $\vec{\sigma}$ the $SU(2)$ matrices, so that S_W becomes

$$\beta\ S_W(U_\ell) = \mathrm{const} - \beta k \cos\theta \ . \tag{46}$$

They then map θ into $r = f(\theta)$ defined by

$$\begin{aligned} Z^{-1} \textstyle\int_0^{r=f(\theta)} x^2\ e^{-x^2/2} dx = \\ (Z')^{-1} \textstyle\int_0^{\theta}\ \sin^2\omega\ e^{\beta k\ \cos\omega} d\omega \ , \end{aligned} \tag{47}$$

with Z, Z' appropriate normalization constants. Defining $\vec{r} = r\hat{r}$, they do an overrelaxed update

$$\begin{aligned} \vec{r}\,' = (1-\omega)\vec{r} + \frac{1}{\sqrt{2}}\, \omega(2-\omega)\vec{\eta} \ , \\ \langle \eta_i \eta_j \rangle_\eta = 2\delta_{ij} \ , \end{aligned} \tag{48}$$

and then remap to get

$$U_\ell' = e^{i\ f^{-1}(r')\hat{r}' \cdot \vec{\sigma}} U_0 \ . \tag{49}$$

This algorithm exactly satisfies detached balance by construction, and so does not need a Metropolis correction step.

Finally, Heller and Neuberger[14] have given a tunable-ω algorithm for the nonlinear σ-model, with action

$$\vec{\phi}^2(x) = 1 \,,$$
$$\beta S = \beta \sum_{x,\hat{\mu}} \vec{\phi}(x) \cdot \vec{\phi}(x + a\hat{\mu}) \,. \tag{50}$$

Defining $\vec{\phi}_0$ to be the value of $\vec{\phi}(x)$ which minimizes S for fixed x, their procedure is to draw a geodesic in the $O(n)$ manifold through $\vec{\phi}$ and $\vec{\phi}_0$ parameterized by ω, thus defining an ω-dependent mean $\bar{\vec{\phi}}'$, and to take $\vec{\phi}'$ distributed around $\bar{\vec{\phi}}'$ with a distribution which approximates the narrowed distribution of eq. (28). In the $O(4)$ case, this recipe can be written in terms of $SU(2)$ matrices as

$$U \equiv \phi^0 + i\,\sigma^a\phi^a \,, \tag{51}$$

$$\bar{U}' = (U_0 U^{-1})^{\omega} U \,,$$
$$U' = V\,\bar{U}' \,, \tag{52a}$$

with V a noise matrix distributed as

$$P(V) = \exp\left\{\frac{\gamma}{2\omega(2-\omega)}\ \mathrm{tr}\ (V)\right\} \,, \tag{52b}$$

where γ is read off from the conditional distribution of U as determined by its nearest neighbors, written as

$$p(U) = \exp\left[\frac{1}{2}\,\gamma\ \mathrm{tr}\ (U\,U_0^{-1})\right] \,. \tag{52c}$$

As before, the new value U' is used as a Metropolis preselection in a final accept/reject step which enforces detailed balance with respect to the action of eq. (50).

5. NUMERICAL RESULTS ON OVERRELAXATION

A number of studies of the utility of overrelaxed algorithms have been carried out in the last two years; we group them by algorithm type as in Sec. 4.

5.1. Studies of microcanonical-analog algorithms

Creutz[11] studied the algorithm OR 0 [cf. eq. (40)] for SU(3) at $\beta = 6.0$ on a $7 \times 7 \times 7 \times 6$ lattice. For the non-gauge-invariant lattice correlation (n_ℓ = number of links)

$$C(U_1,\,U_2) = \frac{1}{n_\ell n} \sum_\ell \mathrm{Re\ Tr}\ \left[U_{\ell 1}^{-1}\,U_{\ell 2}\right] \,, \tag{53}$$

he found that OR 0 outperformed 128 hit Metropolis (which is effectively equivalent to heat bath). For the gauge-invariant plaquette correlation (n_P = number of plaquettes)

$$C_P(U_1,\,U_2) = \frac{1}{n_P} \sum_P \left(W_P - \langle W\rangle\right)_1 \left(W_P - \langle W\rangle\right)_2 \,,$$
$$W_P = \frac{1}{n}\ \mathrm{Re\ Tr}\ U_P \,, \tag{54}$$

OR 0 was somewhat slower at decorrelating plaquettes than heat bath or 10-hit Metropolis. Creutz showed that the plaquette was improved by using an algorithm with a tunable parameter intermediate between OR 0 and standard Metropolis, but he did not study OR n.

Brown and Woch[10] assessed OR n by studying autocorrelation times for the average action and for the "Polyakov loop"[15] defined by

$$\text{“Polyakov”} = \mathrm{Tr}\left[\prod_{\text{line through lattice}} U_\ell\right] \,. \tag{55}$$

Letting $\rho(\Delta)$ be the autocorrelation between sweeps i and $i + \Delta$, they defined the truncated sum

$$\bar{\rho}_m = \sum_{\Delta=-m}^{m} \rho(\Delta) \,, \tag{56}$$

in terms of which the autocorrelation time τ of eq. (2) is given by

$$\tau = \frac{1}{2}\,\bar{\rho}_\infty \,; \tag{57}$$

in practice they estimated $\bar{\rho}_\infty$ from $\bar{\rho}_{150}$. They found that overrelaxation gives a dramatic improvement in decorrelation for $SU(3)$ for $\beta = 5.6$ on a 4^4 lattice, with up to a factor of 3 improvement for interesting observables. The Polyakov loop was best for OR 0, while the action decorrelated best for OR 10 but was worse for OR 0 or OR 1 than for conventional 10-hit Metropolis, consistent with Creutz's results.

Table 1: Correlation time τ in sweeps (error $\sim$ 20).

	OR 1		20 hit Metropolis		Pseudo Heat Bath	
β	R	T	R	T	R	T
6.5	40	90	90	200	130	170
7.0	40	60	110	180		
7.25	40	60	100	170		
7.5	40	60	140	210		

Gupta et al.[16] compared the following algorithms for $SU(3)$:

- 20 hit Metropolis,
- OR n,
- pseudo heat bath (the Cabibbo-Marinari $SU(2)$ subgroup algorithm),
- hybrid Monte Carlo. (58)

They measured small Wilson loops on blocked lattices, blocking in five levels by $\sqrt{3}$/level starting from a 9^4 lattice, and measuring autocorrelations of the blocked loops at each level. Their results for rectangular loops (R) and twisted loops (T) are in Table 1. Gupta et al. conclude that OR 1 is the optimal algorithm for pure gauge simulations over the entire range of accessible β.

In another, independent investigation, Gupta et al.[12] studied the XY model using the algorithm of eq. (42). For an iteration consisting of 15 steps of eq. (42) followed by 2 standard Metropolis steps, they found a sizable reduction in critical slowing down, measuring

$$\tau = 0.15\xi^{1.2} \ . \tag{59a}$$

For comparison, a pure standard Metropolis iteration gives

$$\tau = 5\xi^2 \ . \tag{59b}$$

5.2. Studies of tunable-ω algorithms

Brown and Woch[10] studied the tunable-ω algorithm of eqs. (45)-(49) for the $SU(2)$ subgroups of $SU(3)$, at $\beta = 5.6$ on a 4^4 lattice. They found that the Polyakov loop decorrelated best at $\omega = 2$, where a factor of 3 gain was realized, while the action decorrelated best at an intermediate ω of 1.25. However, more recent work shows that on the Columbia 64-mode machine,[17] a mixed algorithm consisting of 9 $\omega = 2$ iterations followed by 1 conventional Cabibbo-Marinari iteration was not noticeably better than a non-overrelaxed algorithm on a $24^3 \times 16$ lattice at $\beta > 6$. (It is not possible, with current software, to implement the tunable algorithm with $\omega < 2$.)

Adler and Bhanot[18] tested Adler's algorithm of eqs. (43)-(44) for $SU(2)$ at β around 2.2. A factor of 2 improvement in plaquette correlation was observed (with sawtoothing) at very small ω, with $\omega_{opt} \sim 1.025$ on 8^4 and ~ 1.05 on 12^4 lattices. For the magnitude of the Polyakov loop, no distinct minimum was observed, but larger ω values were found to be better than $\omega = 1$, consistent with the Polyakov results of Brown and Woch.

Heller and Neuberger[14] studied the nonlinear σ-model in 1 dimension using the algorithm of eqs. (50)-(52). They point out that interactions can renormalize the starting value of ω into an effective value ω_{eff}. They determine ω_{eff} computationally by studying the field-field autocorrelation matrix and fitting to free-field formulas for the overrelaxation of $\frac{1}{2}\left[(\partial_x \vec{\phi})^2 + m^2 \vec{\phi}^2\right]$, with m the mass gap, getting $\omega_{eff} = \omega - 0.77m$, as compared with the theoretical formula $\omega_{opt}(m) = 2 - 2m$. These formulas show that there exists an $\omega < 2$ for which $\omega_{eff}(\omega, m) \geq \omega_{opt}(m)$, and hence (in $d = 1$ at least) a dynamical critical exponent of $z = 1$ is attainable.

5.3. Conclusions from the numerical work

The following conclusions can be drawn from the numerical work synopsized above:

1. Overrelaxation works! It is the local algorithm of choice in most pure gauge system applications, and for many allied lattice spin and field theories as well.
2. $z = 1$ is not a free field artifact, but is attainable in interacting systems.
3. Sawtoothing of some sort is desirable; in the microcanonical analog versions, one should use OR 1 at least.
4. The decorrelation for the plaquette is best for $1 < \omega_{opt} < 2$, while the Polyakov loop is most improved for $\omega \sim 2$. However, the systematics of (a) which form of overrelaxed algorithm is best, (b) how to pick ω, and (c) which measurements are helped and why, is far from settled—more theoretical and empirical work is needed.

6. FURTHER QUESTIONS AND DISCUSSION

6.1. How does overrelaxation work?

Neuberger[19] has proposed an interesting field-theoretic model for overrelaxation, based on the fact that his "red-black" analysis of the free field case reduces to a 2×2 matrix problem. Let $\phi(x,t)$ be a free field, governed by an action which is the discretization of $\frac{1}{2}[(\partial_\mu\phi)^2 + m^2\phi^2]$. Let $\psi^{\pm}$ be the fields with support at the "red" and "black" sites,

$$\psi^{\pm} = \frac{1}{2}\left[1 \pm (-1)^{\Sigma_\mu x_\mu}\right]\phi \tag{60}$$

and let $\phi^1 = \psi^+ + \psi^- = \phi$ and $\phi^2 = \psi^+ - \psi^- = (-1)^{\Sigma_\mu x_\mu}\phi$ be respectively the sum and difference of the red and black site fields. Neuberger shows that the overrelaxation analysis for ϕ can be mapped into a stochastic field model for the fields $\phi^{1,2}$, with action

$$S = \int d^d x \, \mathcal{L} ,$$

$$\mathcal{L} = \frac{1}{2}\left[(\partial_\mu \phi^1)^2 + (\partial_\mu \phi^2)^2\right] + \frac{1}{2}\left[m_1^2(\phi^1)^2 + m_2^2(\phi^2)^2\right] , \qquad (61)$$

and evolving according to the stochastic differential equation

$$\dot{\phi}^\alpha = -(1-\gamma^2)^{1/2} \frac{\partial S}{\partial \phi^\alpha} - \gamma\, \varepsilon^{\alpha\beta} \frac{\partial S}{\partial \phi^\beta} + (1-\gamma^2)^{1/4}\eta_\alpha ,$$

$$\alpha, \beta = 1, 2,$$

$$\langle \eta_\alpha(x,t)\, \eta_\beta(x',t')\rangle_\eta = 2\delta_{\alpha\beta}\delta^d(x-x')\, \delta(t-t') . \qquad (62)$$

When $\gamma = 0$ eq. (62) is the usual Langevin equation, which corresponds[13] to the extreme underrelaxation limit $\omega = 0$; for $0 < \gamma < 1$ eq. (62) differs from the Langevin equation by the addition of a "reversible mode coupling" term $\varepsilon^{\alpha\beta}\partial S/\partial\phi^\beta$, which enters the Fokker-Planck equation as $(\partial/\partial\phi^\alpha)(\varepsilon^{\alpha\beta}\partial S/\partial\phi^\beta) = 0$ and so leaves $\exp(-\beta S)$ as the equilibrium distribution. According to Neuberger, the model of eqs. (61)-(62) maps into the overrelaxation analysis as follows:

$$m_1^2 = m^2 \ll 1 = \text{light field mass},$$

$$m_2^2 = 4d = \text{heavy field mass}, \qquad (63)$$

$$\gamma = \frac{\omega^2/4}{(1-\omega/2)^2 + \omega^2/4} ,$$

with $0 < \gamma < 1$ mapping into $0 < \omega < 2$. His conclusion is that overrelaxation works by adding a reversible coupling between pairs of long and short wavelength modes. The effect of the reversible coupling is to speed up the stochastic evolution of the long wavelength modes while slowing down the evolution of the short wavelength modes; the fact that the heavy modes are slowed down presumably explains why the observed results of overrelaxation depend on the operator being measured. Based on his analysis, Neuberger suggests that there is a dynamical equivalence class of overrelaxed algorithms which will be tunable to $z = 1$.

6.2. Two questions about overrelaxed Metropolis

An important issue to be answered is whether OR n, or its generalization OR $n_1 n_2$, defined as n_1 steps of the Brown and Woch, Creutz algorithm followed by n_2 steps of standard Metropolis, can attain $z = 1$ for appropriate n or appropriate n_1, n_2. Or do these algorithms merely give an approximation to a fully tunable ω for small lattices?

A second, related question is whether a fully tunable ω can be introduced into overrelaxed Metropolis by an extension of the Brown and Woch, Creutz construction. An interesting possibility is to use a new Metropolis variant proposed by Creutz, Gausterer and Sanielevici,[20] based on unitary link matrices U_i and auxillary "noise" matrices V_i, with U_i, $V_i \varepsilon SU(n)$, iterated according to

$$U_i' = U_i F_i[\{U\}]\, V_i ,$$
$$(V_i')^{-1} = F_i[\{U\}]\, V_i F_i[\{U'\}] , \qquad (64)$$

and with the primed update accepted with conditional probability

$$P_A = \min[1, \exp\beta(-H' + H)] ,$$
$$H = S(\{U\}) + \textstyle\sum_i \tilde{S}(V_i) . \qquad (65)$$

Here $F_i \varepsilon SU(n)$ are arbitrary functionals of the $\{U\}$, and $\exp\left[-\beta\tilde{S}(V)\right]$ is an arbitrary imposed distribution for V. In this generalized setting, OR 0 corresponds to the choices $V_i \equiv 1$ (no noise), $F_i[\{U\}] = U_i^{-1}U_{i0}U_i^{-1}U_{i0}$, so that $U_i F_i[\{U\}] = U_{i0}U_i^{-1}U_{i0}$. An interesting question is whether one can get an efficient tunable-ω algorithm along the lines of the Heller-Neuberger construction of eqs. (51)-(52), by choosing

$$F_i[\{U\}] = \left[U_i^{-1}U_{i0}\right]^\omega , \qquad 0 < \omega < 2 , \qquad (66)$$

and making an appropriate choice of $\tilde{S}(V)$.

6.3. Can $z = 0$ be attained by nonlocal algorithms?

So far we have discussed local algorithms. We now address the question of whether non-local algorithms can give a further improvement in critical slowing down from $z = 1$ to $z = 0$. Two methods have been discussed in the literature: (a) Fourier acceleration in the Langevin equation[21] and (b) stochastic multigrid.[2] In Fourier accelerated Langevin, one performs a Langevin update

$$\phi(x, \tau_{n+1}) = \phi(x, \tau_n) - f_x[\phi, \eta] , \qquad (67)$$

with the driving term f at x coupled nonlocally to the action variation $\delta S/\delta\phi$ at y,

$$f_x = \textstyle\sum_y \left[\varepsilon_{x,y} \frac{\delta S}{\delta\phi(y,\tau_n)} + \sqrt{\varepsilon_{x,y}}\, \eta(y, \tau_n)\right] ,$$
$$\varepsilon_{x,y} = \textstyle\sum_p e^{i\,p\cdot(x-y)}\, \varepsilon(p) . \qquad (68)$$

Usually one takes $\varepsilon(p) = \bar{\varepsilon}(p^2 + m^2)^{-1}$, as motivated by the analysis of critical slowing down in the free field case. In stochastic multigrid, one performs sweeps on successively coarser grids according to the following recursive procedure:

- Do m_1 heat bath sweeps on a L^d lattice;
- Compute the conditional action[2] on the next coarser $(L/2)^d$ lattice;
- Do γ multigrid iterations on the $(L/2)^d$ lattice; and
- Add the coarse grid result back on the L^d lattice and do m_2 further heat bath sweeps.

According to this definition, for each initial sweep on the L^d lattice there are γ on the lattice of dimension $(L/2)^d$, γ^2 on the lattice of dimension $(L/4)^d$, and so forth.

Let me now briefly discuss a number of problems which will have to be surmounted in order to apply these methods as lattice gauge algorithms:

1. For both Fourier acceleration and multigrid—the problem of non-smooth approximate zero modes.[22] We have seen in sect. 3.1 that the worst errors typically come from the smallest eigenvalues of $\bar{L}$. Both Fourier acceleration and multigrid assume that the dangerous modes are localized around the origin in Fourier space (as they are in the free field case): Fourier acceleration through the explicit choice of $\varepsilon_{x,y}$, and multigrid through the blocking scheme. However, as is well-known, non-Abelian theories can have zero eigenmodes which are not localized around zero in Fourier space—in fact, as exemplified by instanton zero modes, they can be arbitrarily rapidly varying. To deal with this problem adaptive versions of the Fourier acceleration or multigrid algorithms[23] will be needed.

2. In Fourier acceleration, the choice of step size is an issue. If the step size is small, the algorithm is accurate but the stochastic evolution is slow. If a large step size is used to get rapid evolution, there are detailed balance errors, or the danger of small acceptances if detailed balance is restored by a global Metropolis step.

3. In multigrid, attention must be paid to the work estimate. If the Lagrangian is a polynomial of order p, the conditional action on level 2^k is computable in $O(p)$ steps from the action on level 2^{k-1}, giving the work estimate[2]
$$W \sim L^d + \gamma(L/2)^d + \gamma^2(L/4)^d + \ldots \tag{69}$$
However, for non-polynomial Lagrangians, such as the Wilson action, the conditional action on level 2^k must in general be computed on the finest lattice and requires L^d steps, giving the work estimate[2]

$$W \sim \left(1+\gamma+\gamma^2+\ldots+\gamma^{\log_2 L}\right) L^d$$
$$= \begin{cases} \sim L^d \log_2 L & \gamma = 1 \\ \sim L^{d+\log_2\gamma} & \gamma > 1 \end{cases} . \tag{70}$$

Now according to Goodman and Sokal,[2] in the Gaussian case one can prove that critical slowing down is completely eliminated for $\gamma \geq 2$. But according to eq. (70), this gives $z_{eff} = \log_2\gamma \geq 1$ and so multigrid does no better than is possible by overrelaxation. Thus, in order for multigrid to beat overrelaxation, one will have to either (i) eliminate slowing down with $\gamma = 1$ (this is conjectured[24] and could be studied numerically in lower dimensional examples even in the absence of proofs) or (ii) find a clever method to reduce the work needed to compute the conditional lattice gauge theory action.

6.4. Conclusion

To conclude, overrelaxation is very simple to implement, and goes half-way towards solving the critical slowing down problem for pure gauge theories—one can get, in principle, from $z = 2$ to $z = 1$. To get from $z = 1$ to $z = 0$ will require further good ideas!

ACKNOWLEDGEMENT

This work was supported by the U.S. Department of Energy under Grant No. DE-AC02-76ER02220.

REFERENCES

1. S.L. Adler, Phys. Rev. Lett. 60, 1243 (1988).
2. J. Goodman and A.D. Sokal, Phys. Rev. Lett. 56, 1015 (1986).
3. S.L. Adler, Phys. Rev. D23, 2901 (1981).
4. S.L. Adler, Phys. Rev. D37, 458 (1988).
5. H. Neuberger, Phys. Rev. Lett. 59, 1877 (1987).
6. See also R.S. Varga, *Matrix Iterative Analysis* (Prentice Hall, Englewood Cliffs, N.J., 1962), Chapt. 4.
7. R.S. Varga, Ref. 6, Chapt. 5; R.W. Hockney and J.W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).

8. S.L. Adler and T. Piran, in *High Energy Physics—1980*, ed. by L. Durand and L.G. Pondrom (AIP Conference Proceedings No. 68, Particles and Fields Subseries No. 22), V.2, p. 958.

9. C. Whitmer, Phys. Rev. **D29**, 306 (1984).

10. F.R. Brown and T.J. Woch, Phys. Rev. Lett. **58**, 2394 (1987).

11. M. Creutz, Phys. Rev. **D36**, 515 (1987).

12. R. Gupta, J. DeLapp, G.G. Batrouni, G.C. Fox, C.F. Baillie, and J. Apostolakis, "The phase transition in the 2-d XY Model," LA-UR-88-2021; C^3P-463.

13. S.L. Adler, Phys. Rev. **D38**, 1349 (1988).

14. U.M. Heller and H. Neuberger, "Overrelaxation and Mode Coupling in Sigma Models," NSF-ITP-88-122.

15. L.D. McLerran and B. Svetitsky, Phys. Lett. **98B**, 195 (1981); J. Kuti, J. Polónyi and K. Szlachányi, Phys. Lett. **98B**, 199 (1981).

16. R. Gupta, G.W. Kilcup, A. Patel, S.R. Sharpe and P. de Forcrand, "Comparison of Update Algorithms for Pure Gauge $SU(3)$," LA-UR-88-824.

17. F.R. Brown, private communication.

18. S.L. Adler and G.V. Bhanot, "Study of an overrelaxation method for gauge theories," IASSNS-HEP-88/34; FSU-SCRI-88/73.

19. H. Neuberger, Phys. Lett. **B207**, 461 (1988).

20. M. Creutz, H. Gausterer and S. Sanielevici, "Biased Monte Carlo Algorithms on Unitary Groups," BNL preprint (August, 1988).

21. G.G. Batrouni, G.R. Katz, A.S. Kronfeld, G.P. Lepage, B. Svetitsky and K.G. Wilson, Phys. Rev. **D32**, 2736 (1985).

22. A. Brandt, S. Solomon and R. Ben-Av, "The Fermionic Matrix, Instantons, Zero-Modes and Multi-Grid," WIS-88/67/OCT-PH.

23. R.G. Edwards, J. Goodman and A.D. Sokal, "Multi-grid Method for the Random Resistor Problem," NYU preprint (March, 1988).

24. A. Sokal, private communication; R. Ben-Av, private communication.

Unpublished; abridged excerpt from full paper archived as: cs.CV/9810017

General Theory of Image Normalization

Stephen L. Adler
Institute for Advanced Study
Princeton, NJ 08540

INTRODUCTION AND BRIEF REVIEW OF VIEWING TRANSFORMATIONS OF A PLANAR OBJECT

A central issue in pattern recognition is the efficient incorporation of invariances with respect to geometric viewing transformations. We focus in this article on a particular method for handling invariances, called "image normalization", which has the capability of extracting all of the invariant features from an image using only a small amount of information about the image (such as a few low order moments). The great appeal of normalization is that it isolates the problem of finding the image modulo the effect of viewing transformations, from the higher order problem of deciding which features of the image are needed for a specific classification decision. Intuitively, normalization is simply a systematic method for transforming from observer–based to image–based coordinates; in the former the image depends on the view, whereas in the latter the image is viewing transformation independent. From a mathematical viewpoint, our method consists of placing a set of constraints on the transformed image equal in number to the number of viewing transformation parameters, permitting one to solve either algebraically or numerically for the parameters of a normalizing transformation. Since the constraints are necessarily viewing transformation noninvariants, their construction is in general simpler than the direct construction of viewing transformation invariants.

Let us begin our discussion with a quick review of the viewing transformations of a planar object, since these transformations will be used as illustrations of our general methods. Under rigid 3D motions the image $I(\vec{x})$, with $\vec{x} = (x_1,\ x_2)$ the two dimensional coordinate in the image plane, is transformed to $I(\vec{x}\,')$, with $\vec{x}\,'$ related to $\vec{x}$ by the *planar projective transformation*

$$x'_n = \frac{\sum_{m=1}^{2} G_{nm}\ x_m + t_n}{1 + \sum_{m=1}^{2} p_m\ x_m},\quad n = 1,2\ . \qquad (1)$$

When the depth of the object is much less than its distance from the lens, then the parameter p_n in Eq. (1) can be neglected, and Eq. (1) reduces to the linear *affine* transformation

$$x'_n = \sum_{m=1}^{2} G_{nm}\ x_m\ + t_n\ . \qquad (2)$$

[An affine transformation, with G_{nm} replaced by $G_{nm} - t_n\ p_m$, also results when Eq. (1) is expanded in a power series in x_m and second and higher order terms are neglected.] Additionally, when the viewed object is constrained to lie in the plane normal to the viewing or 3 axis, Eq. (2) specializes further to the *similarity* transformation group of scalings, rotations, and translations, in which G_{nm} is simply a multiple (the scale factor) of a two dimensional rotation matrix. The projective transformations, the affine transformations, and the similarity transformations all form groups, and this will be the characterizing feature of the viewing transformations studied in our general analysis.

In applications, it will be convenient to use subgroup factorizations, which are readily obtained from the group multiplication rule for the transformations of Eqs. (1) and (2). For example, a general planar projective transformation can be written as the result of composing what we will term a *restricted projective transformation*

$$x''_n = \frac{x'_n}{1 + \sum_{m=1}^{2} p_m\ x'_m} \qquad (3)$$

with the general affine transformation of Eq. (2). Another subgroup factorization expresses the general affine transformation of Eq. (2) as the result of the composition of a pure translation

$$x''_n = x'_n + t_n \qquad (4a)$$

with a homogeneous affine transformation

$$x'_n = \sum_{m=1}^{2} G_{nm}\ x_m\ . \qquad (4b)$$

Yet a third subgroup factorization expresses a general homogeneous affine transformation as the result of composing what we will term a *restricted affine transformation*, which has vanishing upper right diagonal matrix element,

$$x''_n = \sum_{m=1}^{2} g_{nm}\ x'_m\ ,\quad g_{12} = 0\ , \qquad (5a)$$

with a pure rotation

$$\begin{aligned} x'_n &= \sum_{m=1}^{2} R_{nm} x_m \ , \\ R_{11} &= R_{22} = \cos\theta, \quad R_{12} = -R_{21} = -\sin\theta \ . \end{aligned} \tag{5b}$$

A variant of Eqs. (5a,b) is obtained by requiring that the matrix g have unit determinant, so that it has the two-parameter form $g_{11} = u$, $g_{12} = 0$, $g_{21} = w$, $g_{22} = u^{-1}$, and then including a scale factor λ in Eq. (5b), which now reads

$$x'_n = \lambda \sum_{m=1}^{2} R_{nm} x_m . \tag{5c}$$

GENERAL THEORY OF IMAGE NORMALIZATION

We proceed now to formulate a general framework for image normalization, with the aim of understanding the common elements of the various normalization methods which appear in the literature and of generalizing them to new applications. As a preliminary to the mathematical discussion of Subsecs. 2A-E, we specify our notation for viewing transformations. Let $\mathcal{G} = \{S\}$ be a group of symmetry or viewing transformations S, which act on the image $I(\vec{x})$ according to

$$I(\vec{x}) \to I_S(\vec{x}) = I(\vec{S}(\vec{x})) \ . \tag{6a}$$

Our notational convention, that we shall adhere to throughout, is that $\vec{x}' = \vec{S}(\vec{x})$ is the concrete image coordinate mapping induced by the abstract group element S. [A specific example of such a transformation would be the planar projection transformation of Eq. (1), in which S would be the abstract element of the planar projective group characterized by the parameters G_{mn}, t_n, p_m specifying the concrete coordinate mapping.] In this notation, the result of successive transformations with S_1 followed by S_2 is given by

$$I(\vec{x}) \to I_{S_2 S_1}(\vec{x}) = I(\vec{S}_2(\vec{S}_1(\vec{x}))) \ . \tag{6b}$$

The transformation groups of interest to us are in general ones with continuous parameters, in other words, Lie groups. However, very little of the formal apparatus of Lie group theory is required in what follows; basically, all we use is the group closure property and the enumeration of the number of group parameters. In particular, no knowledge of the representation theory of Lie groups is needed.

A. *The normalization recipe.* We begin by giving the general prescription for an image normalization transformation. Let $\vec{N}_I(\vec{x})$ be a transformation of $\vec{x}$ which depends on the image I, and which is constructed so that under the image transformation of Eq. (6a), it behaves as

$$\vec{N}_{I_S}(\vec{x}) = \vec{S}^{-1}(\vec{N}_I(\vec{x})) \ , \tag{7a}$$

with $\vec{S}^{-1}$ the inverse transformation to $\vec{S}$ of Eq. (6a),

$$\vec{S}(\vec{S}^{-1}(\vec{x})) = \vec{x} \ . \tag{7b}$$

Also, let $\vec{M}_I(\vec{x})$ be an optional second transformation of $\vec{x}$ which depends on the image I only through invariants under the group of transformations $\mathcal{G}$, that is,

$$\vec{M}_{I_S}(\vec{x}) = \vec{M}_I(\vec{x}), \qquad \text{all } S \in \mathcal{G} \ . \tag{7c}$$

Then

$$\tilde{I}(\vec{x}) = I(\vec{N}_I(\vec{M}_I(\vec{x}))) \tag{8}$$

is a normalized image which is invariant under all transformations of the group $\mathcal{G}$. This is an immediate consequence of Eq. (6a) and Eqs. (7a–c), from which we have

$$\begin{aligned} \tilde{I}_S(\vec{x}) &= I_S(\vec{N}_{I_S}(\vec{M}_{I_S}(\vec{x}))) \\ &= I(\vec{S}(\vec{S}^{-1}(\vec{N}_I(\vec{M}_I(\vec{x}))))) \\ &= I(\vec{N}_I(\vec{M}_I(\vec{x}))) = \tilde{I}(\vec{x}) \ . \end{aligned} \tag{9}$$

B. *Uniqueness.* Before specifying how to actually construct a map $\vec{N}_I$ obeying Eq. (7a), let us address the issue of uniqueness. That is, given *two* maps $\vec{N}_{1I}(\vec{x})$ and $\vec{N}_{2I}(\vec{x})$, both of which obey Eq. (7a), how are they related? By hypothesis, we have

$$\begin{aligned} \vec{N}_{1I_S}(\vec{x}) &= \vec{S}^{-1}(\vec{N}_{1I}(\vec{x})) \ , \\ \vec{N}_{2I_S}(\vec{x}) &= \vec{S}^{-1}(\vec{N}_{2I}(\vec{x})) \ . \end{aligned} \tag{10}$$

Since for any $\vec{f}(\vec{x})$ and $\vec{g}(\vec{x})$ we have

$$\vec{f}(\vec{g}(\vec{x}))^{-1} = \vec{g}^{\,-1}(\vec{f}^{-1}(\vec{x})) \ , \tag{11a}$$

we can rewrite the first line of Eq. (10) as

$$\vec{N}_{1I_S}^{-1}(\vec{x}) = \vec{N}_{1I}^{-1}(\vec{S}(\vec{x})) \ . \tag{11b}$$

Let us now define a new map $\vec{M}_I(\vec{x})$ by

$$\vec{M}_I(\vec{x}) \equiv \vec{N}_{1I}^{-1}(\vec{N}_{2I}(\vec{x})) \ , \tag{12}$$

which reduces to the identity map when $\vec{N}_{1I} = \vec{N}_{2I}$; then by Eq. (11b) and the first line of Eq. (10), we have

$$\begin{aligned} \vec{M}_{I_S}(\vec{x}) &= \vec{N}_{1I_S}^{-1}(\vec{N}_{2I_S}(\vec{x})) \\ &= \vec{N}_{1I}^{-1}(\vec{S}(\vec{S}^{-1}(\vec{N}_{2I}(\vec{x})))) \\ &= \vec{N}_{1I}^{-1}(\vec{N}_{2I}(\vec{x})) = M_I(\vec{x}) \ . \end{aligned} \tag{13a}$$

In other words, $M_I(\vec{x})$ depends on the image I only through invariants under transformations of the group $\mathcal{G}$,

and from Eq. (12), the normalizing map $\vec{N}_{2I}$ is related to the normalizing map $\vec{N}_{1I}$ by

$$\vec{N}_{2I}(\vec{x}) = \vec{N}_{1I}(\vec{M}_I(\vec{x})) \quad . \tag{13b}$$

This is why in writing the general normalized image corresponding to a particular normalizing map in Eq. (8), we have included in the $\vec{x}$ dependence the possible appearance of a map $\vec{M}_I$ which depends on the image only through invariants under transformation by elements of $\mathcal{G}$.

C. *Construction of $\vec{N}_I$ by imposing constraints, and demonstration that normalization yields a complete set of invariants.* We next show that one can construct an image normalization transformation obeying Eq. (7a) by imposing a suitable set of constraints. We shall assume now that $\mathcal{G}$ is a K–parameter Lie group which is continuously connected to the identity. Let $C_k[I] = C_k[I(\vec{x})]$, $k = 1, ..., K$ (where $\vec{x}$ is a dummy variable) be a set of functionals of the image $I(\vec{x})$ with the property that the K constraints

$$C_k[I_{S'}] = C_k[I(\vec{S}'(\vec{x}))] = 0 \ , \quad k = 1, ..., K \tag{14a}$$

are satisfied for a unique element $S' = N_I$ of $\mathcal{G}$, so that

$$C_k[I(\vec{N}_I(\vec{x}))] = 0 \ , \quad k = 1, ..., K \ . \tag{14b}$$

Then, as we shall now show, $\vec{N}_I(\vec{x})$ is the desired normalizing transformation.

We remark that the condition that Eqs. (14a,b) should have a unique solution can be relaxed in applications to the condition that there be only one solution in the range of relevant viewing transformation parameters. Clearly, either form of the uniqueness condition requires that the constraint functionals *not* be invariants under $\mathcal{G}$, and thus their structure will in general be simpler than that of directly constructed viewing transformation invariants. In many cases, the constraints can be constructed from viewing transformation *covariants*, which have simple algebraic properties under the transformations of $\mathcal{G}$, permitting closed form algebraic solution for the parameters of the normalizing transformation.

To see that the construction of Eqs. (14a,b) gives a transformation $\vec{N}_I(\vec{x})$ that obeys Eq. (7a), let us consider the effect of replacing I by I_S in Eqs. (14a,b). By hypothesis, the constraints

$$C_k[I_S(\vec{S}'(\vec{x}))] = 0 \ , \quad k = 1, ..., K \tag{15a}$$

are uniquely satisfied by a group element $S' = N_{I_S}$ of $\mathcal{G}$, so that

$$C_k[I_S(\vec{N}_{I_S}(\vec{x}))] = 0 \ , \quad k = 1, ..., K \ , \tag{15b}$$

with $\vec{N}_{I_S}(\vec{x})$ the proposed normalizing transformation corresponding to I_S. But using Eq. (6a), we can also write Eq. (15b) as

$$C_k[I(\vec{S}(\vec{N}_{I_S}(\vec{x})))] = 0 \ , \quad k = 1, ..., K \ , \tag{15c}$$

which has the same structure as Eq. (14b). Therefore, by uniqueness of the solution N_I of Eq. (14b) we must have

$$\vec{S}(\vec{N}_{I_S}(\vec{x})) = \vec{N}_I(\vec{x}) \ , \tag{16a}$$

which by Eq. (7b) is equivalent to

$$\vec{N}_{I_S}(\vec{x}) = \vec{S}^{-1}(\vec{N}_I(\vec{x})) \ , \tag{16b}$$

showing that the N_I produced by solving the constraints does indeed obey Eq. (7a). Hence the imposition of constraints gives a constructive procedure for generating image normalization transformations.

We note that this construction makes the normalizing transformation $\vec{N}_I$ an element of the group $\mathcal{G}$, and the quotient $\vec{M}_I(\vec{x}) \equiv \vec{N}_{1I}^{-1}(\vec{N}_{2I}(\vec{x}))$ of two normalizing maps constructed by imposing different sets of constraints will likewise be an element of $\mathcal{G}$. When both $\vec{N}_I$ and $\vec{M}_I$ in Eq. (8) belong to $\mathcal{G}$, we can invert Eq. (8) to express the original image I in terms of the invariant, normalized image $\tilde{I}$ according to

$$I(\vec{x}) = \tilde{I}(\vec{M}_I^{-1}(\vec{N}_I^{-1}(\vec{x}))) \ . \tag{16c}$$

This equation shows that normalization leads to a complete set of invariants, in the sense that the information in the normalized image, plus the K parameters determining the viewing transformation $\vec{M}_I^{-1}(\vec{N}_I^{-1}(\vec{x}))$, suffice to completely reconstruct the original image. By way of contrast, representation–theoretic and integral transform methods, although attacking the same problem as is discussed here, yield only a small fraction of the complete set of invariants. Moreover, normalization has the further advantage of requiring only a minimal knowledge of the kinematic structure of the group; the full irreducible representation structure is not needed, and the methods described here are applicable to noncompact as well as to compact groups. We note finally that the discussion of this section is slightly less general than that of Subsections A,B above, where we did not require either $\vec{N}_I$ or $\vec{M}_I$ to belong to $\mathcal{G}$; the most general normalizing map $\vec{N}_I$ is obtained from one generated by constraints by using as its argument a map $\vec{M}_I$ which does not belong to $\mathcal{G}$ but that is invariant under transformations of the image I by $\mathcal{G}$.

D. *Extension to reflections and contrast invariance.* We consider next two simple extensions of the constraint method for constructing the normalizing transformation. The first involves relaxing the requirement that $\mathcal{G}$ be simply connected to the identity, as is needed if $\mathcal{G}$ contains improper transformations such as reflections. Reflections are said to be independent if they do not differ solely by an element of the connected component of the group; for each independent discrete reflection R in $\mathcal{G}$, the set of constraints of Eq. (14a) must be augmented by an additional constraint $D[I(\vec{S}'(\vec{x}))] > 0$, where $D[I(\vec{x})]$ is a

functional of the image which changes sign under the reflection operation R,

$$D[I(\vec{R}(\vec{x}))] = -D[I(\vec{x})] \ . \tag{17}$$

The second extension involves incorporating invariance under changes of image contrast, that is, under image transformations of the form

$$I(\vec{x}) \to cI(\vec{x}), \qquad c > 0 \ . \tag{18a}$$

To the extent that illumination is sufficiently slowly varying that it can be treated as constant over a viewed object, changes in illumination level as the object is moved to different views take the form of changes in the constant c in Eq. (18a), which is why incorporating contrast invariance can be important. If we require that the constraint functionals C_k [and D if needed] should be invariant under the change of contrast of Eq. (18a), then the image normalization transformation $\vec{N}_I(\vec{x})$ and the auxiliary transformation $\vec{M}_I(\vec{x})$ can be taken to be contrast invariant. A contrast invariant normalized image $\tilde{I}_c(\vec{x})$ is then obtained by the obvious recipe

$$\tilde{I}_c(\vec{x}) = \frac{\tilde{I}(\vec{x})}{\int d^2x \tilde{I}(\vec{x})} \ . \tag{18b}$$

E. *Use of subgroup decompositions.* Suppose that for a general element S of the group $\mathcal{G}$, there is a subgroup decomposition of the form

$$S = S_2 S_1 \ , \tag{19a}$$

with S_2 belonging to a subgroup $\mathcal{G}_2$ of $\mathcal{G}$, S_1 belonging to a subgroup $\mathcal{G}_1$ of $\mathcal{G}$, and with the respective parameter counts K, K_1, and K_2 of $\mathcal{G}$, $\mathcal{G}_1$, and $\mathcal{G}_2$ obeying

$$K = K_1 + K_2 \ . \tag{19b}$$

(Such subgroup compositions for a general Lie group are obtained by constructing a composition series for the group, but we will not need this formal apparatus in the relatively simple applications that follow.) Let us suppose further that we can solve the problem of image normalization with respect to the group $\mathcal{G}_1$, and that we wish to extend this solution to the full invariance group $\mathcal{G}$. The subgroup decomposition allows this to be done by imposing K_2 additional constraints to deal with the $\mathcal{G}_2$ subgroup, as follows. Let $C_{2k}[I(\vec{x})]$, with $k = 1, ..., K_2$, be a set of functionals of the image chosen so that the constraints

$$C_{2k}[I(\vec{N}_{2I}(\vec{S}_1(\vec{x})))] = 0 \ , \quad k = 1, ..., K_2 \tag{20a}$$

are independent of $S_1 \in \mathcal{G}_1$. In particular, taking S_1 as the identity transformation, Eq. (20a) simplifies to

$$C_{2k}[I(\vec{N}_{2I}(\vec{x}))] = 0 \ , \quad k = 1, ..., K_2 \ , \tag{20b}$$

which if we impose the requirement of a unique solution over transformations $N_2 \in \mathcal{G}_2$ determines a "partial normalization" transformation $\vec{N}_{2I}$. Note that a sufficient condition for the constraints of Eq. (20a) to be independent of S_1 is for the functionals C_{2k} to be S_1-independent, but this is not a necessary condition; we will see examples in which, as S_1 traverses $\mathcal{G}_1$, the functionals are merely covariant in some simple way that guarantees invariance of the constraints obtained by equating all the functionals to zero. To see how $\vec{N}_{2I}$ transforms under the action of the group $\mathcal{G}$, we replace I by I_S in Eq. (20b), giving

$$C_{2k}[I_S(\vec{N}_{2I_S}(\vec{x}))] = 0 \ , \quad k = 1, ..., K_2 \quad ; \tag{21a}$$

again making use of Eq. (6a) this becomes

$$C_{2k}[I(\vec{S}(\vec{N}_{2I_S}(\vec{x})))] = 0 \ , \quad k = 1, ..., K_2 \quad . \tag{21b}$$

Since the argument $\vec{S}(\vec{N}_{2I_S}(\vec{x}))$ appearing in Eq. (21b) is no longer a member of the $\mathcal{G}_2$ subgroup, we cannot conclude that it is equal to the argument $\vec{N}_{2I}(\vec{x})$ appearing in Eq. (20b), but the arguments can differ at most by a transformation of $\vec{x}$ by some member $\vec{S}_1'$ of the subgroup $\mathcal{G}_1$ which leaves the constraints invariant, giving

$$\vec{N}_{2I_S}(\vec{x}) = \vec{S}^{-1}(\vec{N}_{2I}(\vec{S}_1'(\vec{x}))) \tag{22a}$$

as the subgroup analog of Eq. (7a). Corresponding to this, the partially normalized image defined by

$$\tilde{I}(\vec{x}) = I(\vec{N}_{2I}(\vec{x})) \tag{22b}$$

transforms under the group $\mathcal{G}$ as

$$\begin{aligned} \tilde{I}(\vec{x}) \to \tilde{I}_S(\vec{x}) &= I_S(\vec{N}_{2I_S}(\vec{x})) \\ &= I(\vec{S}(\vec{S}^{-1}(\vec{N}_{2I}(\vec{S}_1'(\vec{x}))))) \\ &= I(\vec{N}_{2I}(\vec{S}_1'(\vec{x}))) = \tilde{I}(\vec{S}_1'(\vec{x})) \ , \end{aligned} \tag{22c}$$

and thus changes only by a transformation lying in the $\mathcal{G}_1$ subgroup. Further image normalization of $\tilde{I}$ using the constraints appropriate to $\mathcal{G}_1$ then gives a final normalized image

$$\hat{I}(\vec{x}) = I(\vec{N}_{2I}(\vec{N}_{1\tilde{I}}(\vec{M}_I(\vec{x})))) \ , \tag{23}$$

which is invariant with respect to the full group of transformations $\mathcal{G}$, where as before $\vec{M}_I$ is any transformation which is constructed solely using $\mathcal{G}$ invariants of the image.

Pergamon

Pattern Recognition, Vol. 31, No. 10, pp. 1551–1556, 1998

SIMILARITY AND AFFINE NORMALIZATION OF PARTIALLY OCCLUDED PLANAR CURVES USING FIRST AND SECOND DERIVATIVES

STEPHEN L. ADLER* and RANGANATHAN KRISHNAN†

Institute for Advanced Study, Princeton, NJ 08540, U.S.A.
†Dragon Systems, 320 Nevada Street, Newton, MA 02169, U.S.A.

(*Received* 12 *May* 1997; *in revised form* 12 *January* 1998)

Abstract—We apply the general framework for image normalization to the problem of the similarity and affine normalization of partially occluded planar curves. An algorithm is given using first derivatives to give a similarity normalization, and first and second derivatives to give an affine normalization, dependent on a finite interval around a chosen point P of the curve. © 1998 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved

Normalization   Affine   Similarity   Partially occluded   Planar curves   Invariants
Affine invariants   Similarity invariants

1. INTRODUCTION, AND NORMALIZATION WITH RESPECT TO TRANSLATIONS AND ROTATIONS

The problem of efficiently incorporating invariances plays an important role in machine vision, and much work has gone into methods for extracting features invariant with respect to classes of geometric viewing transformations. An appealing way to deal with this problem is to use image normalization, in which the transformed image is reduced to a standard form which effectively "mods out" the viewing transformation group; the normalized image and all of its features are then invariants, while the parameters of the normalizing transformation give the pose of the original image, so there is no loss of information. Adler[1] recently gave a general formal framework for image normalization, and illustrated it by using it to rederive and generalize known methods for the normalization of a non-occluded, isolated image. The method is not, however, limited to this rather special case, as we demonstrate here by applying it to the more realistic problem of the similarity and affine normalization of a partially occluded planar curve, such as that characterizing the image of the boundary of a partially occluded planar object.

There has been much discussion in the literature of the problem of the recognition of partially occluded curves, using mainly the approach of constructing viewing transformation *invariants* using strictly local information at a point P (obtained by computing a finite number of derivatives; see reference (2) and references cited therein), or as in recent work of Bruckstein *et al.*[3] using "semi-local" information constructed from a finite neighborhood of P. Our approach is similar in spirit to this latter work, but instead of directly constructing invariants we instead use viewing transformation *covariants*, that is quantities which are not invariant but have simple algebraic transformation properties under the transformations of interest, to construct a normalization algorithm. The advantage of focusing on covariants is that they often can be constructed in a more computationally robust fashion (for example, using only low-order derivatives or moments) than is possible for invariants.

Let us assume that we are given a set of closed planar template curves, $C_1, C_2, \ldots, C_N$, with C_j specified by giving its parametric form $\mathbf{x}_j(t_j) = [x_j(t_j), y_j(t_j)]$ in which the parameter t_j increases as the curve C_j is traversed counterclockwise. We are now given a convex target curve segment $\mathbf{x}_{\text{target}}(t_{\text{target}}) = [x_{\text{target}}(t_{\text{target}}), y_{\text{target}}(t_{\text{target}})]$ for $t_{\text{target min}} \le t_{\text{target}} \le t_{\text{target max}}$, which represents a fragment of the jth template as distorted by both an affine viewing transformation and a possible reparameterization $t_{\text{target}} = U(t_j)$ with unknown function U. The problem is to identify the fragment with the correct corresponding segment of the correct template. In the discussion that follows, we shall omit the subscripts j and "target" from the curve parameters, which will simply be referred to by the generic label t; the fact that t has a different meaning for each of the different curves will be taken into account by framing the entire algorithm in terms of reparameterization invariant quantities.

We proceed by using the subgroup method (see reference (1), and Appendix A of this paper, where the methods of reference (1) are adapted to the normalization of planar curves), in which we successively solve the problem for translations and rotations, for the full

*Author to whom correspondence should be addressed.

similarity group, and finally for the full affine group. For translations and rotations the procedure is obvious. One first associates with each point P of the target and templates the reparameterization invariant unit normalized tangent vector

$$\hat{T} = \mathbf{T}/|\mathbf{T}|, \tag{1a}$$

where the unnormalized tangent vector is given by

$$\mathbf{T} = (x', y'), \tag{1b}$$

and where the prime denotes differentiation with respect to the generic parameter t. One first takes a particular point P of the target curve, translates it to the origin, and rotates the target so that it is tangent to the y axis at the origin and has its tangent vector pointing into the upper half-plane. One next picks a point P' on the first template, translates it to the origin, and rotates the first template so that it is also tangent to the y-axis at the origin and has its tangent vector pointing into the upper half-plane. Maintaining these conditions, one sweeps P' over the first template and looks for a match, proceeding in this fashion from template to template until a match is found. If no match is found, one then repeats the procedure with the tangent vectors of target and template at the origin pointing in opposite directions. The search implicit in this procedure is necessary because, without identifying landmarks on the curves, there is no way of knowing *a priori* (i) which point P' of the correct template corresponds to the given point P on the target, and (ii) whether the parameter of the target runs in the same or the opposite sense to that of the template. Apart from the search over the possibilities for P' and the relative sense of the template and target parameters, the algorithm consists simply of imposing the same translational and rotational normalization on the template and target curves.

2. NORMALIZATION WITH RESPECT TO SIMILARITY TRANSFORMATIONS

We turn next to the problem of normalization with respect to the full similarity transformation group, comprising translations, rotations, and scalings. Let us pick a specific point P of the target, a neighborhood of which will be used to construct the similarity normalization. As before, we perform a translational and rotational normalization by translating P to the origin and rotating the target so that $x' = 0$, $y' > 0$ at P. We begin by observing that the tangent vector to the target at a general point with parameter t makes an angle with respect to the y-axis given by

$$w(t) \equiv \tan\theta = \frac{x'}{y'}, \tag{2}$$

which is reparameterization invariant, and which by our translational and rotational normalization vanishes at P. Let us compute $w(t)$ for each t on the target segment and store its value along with $x(t)$, $y(t)$. Let now $d\Gamma(w) \geq 0$ with $\Gamma(0) = 0$ be a measure for integration over the target curve; because w is invariant under the coordinate rescaling $x \to \lambda x$, $y \to \lambda y$, so is this measure. Also let $\Delta_S > 0$ be a parameter governing the size of an integration interval along the target segment starting from P, assumed to lie entirely within the segment. Finally, let $F_D(x, y)$ be any nonnegative function of x, y which is homogeneous of degree D under coordinate rescaling, that is

$$F_D(\lambda x, \lambda y) = \lambda^D F_D(x, y). \tag{3}$$

We now use the quantities just defined to form the constraint (with $\mu \neq \nu$ arbitrary real number parameters)

$$\frac{\int_0^{\Delta_s} d\Gamma(w) F_D(\bar{\lambda}x(t), \bar{\lambda}y(t))^\mu}{\int_0^{\Delta_s} d\Gamma(w) F_D(\bar{\lambda}x(t), \bar{\lambda}y(t))^\nu} = 1, \tag{4}$$

which using the homogeneity of F_D can be solved algebraically for the scale parameter $\bar{\lambda}$ to give

$$\bar{\lambda} = R^{-1/[D(\mu-\nu)]},$$

$$R = \frac{\int_0^{\Delta_s} d\Gamma(w) F_D(x(t), y(t))^\mu}{\int_0^{\Delta_s} d\Gamma(w) F_D(x(t), y(t))^\nu}. \tag{5}$$

Then according to the general normalization prescription of reference (1) and Appendix A, the normalized curve $\tilde{C}$ with the parameterized form $\tilde{\mathbf{x}}(t) = [\bar{\lambda}x(t), \bar{\lambda}y(t)]$ is invariant with respect to scaling, as well as to translation and rotation, of the target segment. Note that the resulting similarity normalization depends not only on the parameter Δ_S, the exponents μ, ν and the chosen functions $\Gamma(w)$ and $F_D(x, y)$, but also on the choice of the fiducial point P, and hence the normalized curves corresponding to different choices of P will not in general be congruent to one another.

To determine the correspondence between target segment and the templates, we now do a search over the templates, over the choice of fiducial point P' on each template, and over the two possible senses of the tangent vector at P', applying the same normalization recipe as was applied to the target segment and looking for a match. The search for a match is done by comparing the normalized target and template using any convenient measure for the degree to which they coincide (such as, e.g. the integral over all points of the normalized target of the minimum distance to the normalized template). In some applications it may suffice to use a measure based on only a small number of features extracted from the normalized curves, in which case an alternative procedure would be to dispense with normalization and directly construct invariant features by the methods of references (2, 3).

3. AFFINE NORMALIZATION

We turn finally to the problem of normalization with respect to the affine transformation group. Again let us pick a specific point P of the target, a neighbor-

hood of which will be used to construct the affine normalization. As before, we perform a translational and rotational normalization by translating P to the origin and rotating the target so that $x' = 0$, $y' > 0$ at P. The conditions that P lie at the origin and that $x' = 0$ at P are preserved only by a subgroup of the full affine group, consisting of homogeneous affine transformations of the form

$$A = \begin{pmatrix} \alpha & 0 \\ \beta & \gamma \end{pmatrix}. \tag{6}$$

Since we have been implicitly dealing only with proper transformations (as opposed to reflections), we assume that the matrix of equation (6) has a positive determinant, and so $\alpha\gamma > 0$.

Under the action of A, the coordinate vector $\mathrm{x} = (x, y)$ is transformed to $\mathrm{x}_A = (x_A, y_A) = (\alpha x, \beta x + \gamma y)$. Let us now follow Vaz and Cyganski[4] and introduce a new arc length parameter $d\tau$ defined by

$$d\tau = |x'y'' - x''y'|^{1/3}\,dt, \tag{7}$$

which is shown in reference (4) to be reparameterization invariant. Under the action of A, $d\tau$ is transformed to $d\tau_A = (\det A)^{1/3}\,d\tau = (\alpha\gamma)^{1/3}\,d\tau$, and so although $d\tau$ is not an affine invariant, it transforms linearly under affine transformations and thus is a convenient integration measure. Clearly, the *difference* between the upper and lower limits of an integration over τ also simply rescales by $(\alpha\gamma)^{1/3}$ under the affine transformation A; henceforth, we shall choose the constant of integration in the integrated version of equation (7) so that $\tau = 0$ at P.

We now define "center of mass" coordinates x_{CM}, y_{CM} and central moment integrals $\mu_{rs} = \mu_{rs}(T)$, in the neighborhood of P, by

$$(x_{CM}, y_{CM}) = \frac{\int_0^T d\tau[x(t), y(t)]}{\int_0^T d\tau} \tag{8a}$$

and

$$\mu_{rs}(T) = \int_0^T d\tau (x - x_{CM})^r (y - y_{CM})^s, \tag{8b}$$

where the upper limit T will be specified through the normalization procedure. Under the action of the affine transformation with matrix A, the central moments transform to

$$\mu_{A;rs}(T_A) = \int_0^{T_A} d\tau_A (x - x_{CM})_A^r (y - y_{CM})_A^s, \tag{8c}$$

with $T_A = (\alpha\gamma)^{1/3} T$. Substituting the transformed quantities and making a change of integration variable from τ_A to τ, we get

$$\begin{aligned} \mu_{A;rs}(T_A) &= (\alpha\gamma)^{1/3} \int_0^T d\tau[\alpha(x - x_{CM})]^r \\ &\quad \times [\beta(x - x_{CM}) + \gamma(y - y_{CM})]^s \\ &= (\alpha\gamma)^{1/3} \sum_{t=0}^{s} \frac{s!}{t!(s-t)!} \alpha^r \beta^t \gamma^{\,s-t} \mu_{r+t\,s-t}(T). \end{aligned} \tag{9}$$

Let us now impose a set of four constraints to uniquely fix the affine transformation parameters α, β, γ and the interval of integration $T = \mu_{00}$,

$$\mu_{A;02}(T_A) = \mu_{A;20}(T_A) = 1, \quad \mu_{A;11}(T_A) = 0, \quad T_A = f, \tag{10}$$

with f a positive constant characterizing the normalization. Because the μ's have been defined as central moments, these conditions are always attainable for f in the range of T_A. Substituting equaton (9), the first three conditions of equation (10) can be solved algebraically in terms of T by solving a quadratic equation. Writing

$$\hat{\gamma} = \frac{\gamma}{\alpha}, \qquad \hat{\beta} = \frac{\beta}{\alpha}, \tag{11a}$$

the solution takes the form

$$\begin{gathered} \hat{\gamma} = \frac{\mu_{20}(T)}{\mathscr{D}^{1/2}}, \quad \hat{\beta} = -\frac{\mu_{11}(T)}{\mathscr{D}^{1/2}}, \\ \mathscr{D} = \mu_{20}(T)\mu_{02}(T) - \mu_{11}(T)^2, \\ \alpha = \mu_{20}(T)^{-3/8}\,\hat{\gamma}^{-1/8}, (\alpha\gamma)^{1/3} = \mathscr{D}^{-1/8}. \end{gathered} \tag{11b}$$

The final condition of equation (10), which determines the integration interval T before normalization, is implicit and must be solved by an iterative method. Writing

$$F(T) = \mathscr{D}^{-1/8}\mu_{00} - f, \tag{12a}$$

the desired value of T is a solution of $F(T) = 0$, which always exists for f in the range of $\mathscr{D}^{-1/8}\mu_{00}$. For simplicity, we solve this equation using Newton's method, for which the desired solution T is the limit T_∞ of the iteration defined by

$$T_{n+1} = T_n - \frac{F(T)}{F'(T)}, \tag{12b}$$

with the prime here denoting differentiation with respect to T, and with the initialization of T_1 specified below. The result of this iteration, when substituted into equations (11a) and (11b), is the set of parameters α, β, γ of a normalizing affine transformation matrix A with the form of equation (6). In order to get a normalizing transformation that behaves smoothly as $T \to 0$, we take as the final normalizing transformation the rescaled matrix $\bar{A} = A_{\text{unit circle}}^{-1} A$, with $A_{\text{unit circle}}$ the transformation obtained by applying the above procedure to a unit circle passing through the origin and tangent to the y-axis there. Corresponding to this choice, we parameterize f as $f = f(\Delta_A)$, with Δ_A the arc length along a unit circle and with

$$f(\Delta_A) = \mathscr{D}_{\text{unit circle}}^{-1/8}(\Delta_A)\Delta_A, \tag{12c}$$

so that the limit of vanishing normalization interval is simply the limit $\Delta_A \to 0$. With this choice of f, a convenient initialization for the Newton iteration is $T_1 = \Delta_A$. (Formulas for the central moment integrals and $\mathscr{D}^{-1/8}$ computed from such a unit circle are given in

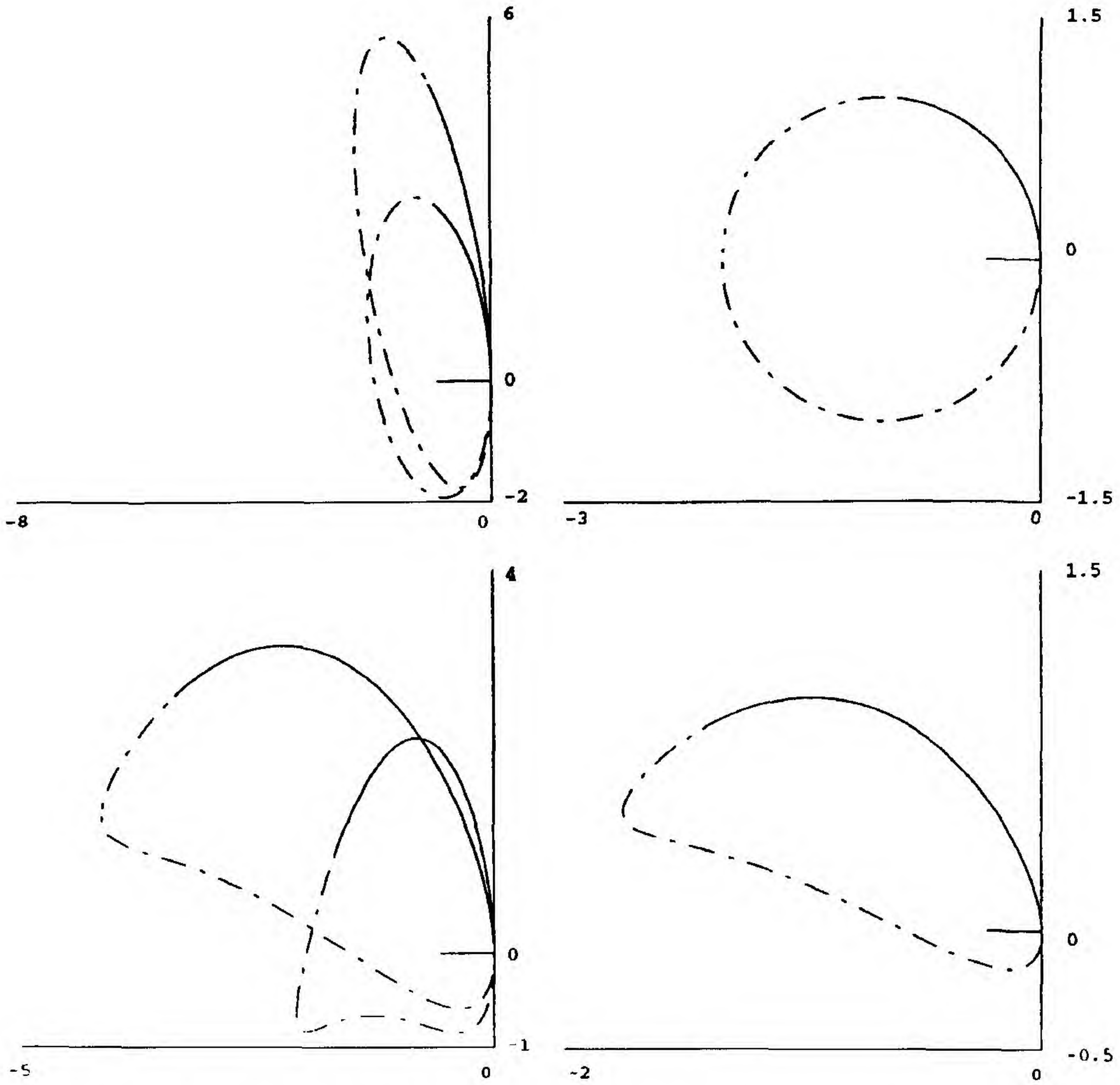


Fig. 1. Normalization results. Ellipses related by affine transformation (upper left) normalize to the same unit circle (upper right). Generic curves related by affine transformation (lower left) normalize to the same curve (lower right). The solid portion of each curve is the segment used for normalization. Typically 5–6 Newton iterations were used.

Appendix B.) The matrix $\tilde{A}$ again has the form of equation (6), with parameters $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$, and the final normalized curve $\tilde{C}$ has the parameterized form

$$\tilde{\mathrm{x}}(t) = \mathrm{x}_{\tilde{A}}(t) = [\tilde{\alpha}x(t), \tilde{\beta}x(t) + \tilde{\gamma}y(t)]. \tag{13}$$

The transformation $\tilde{A}$ has the useful property that it transforms any ellipse C through the origin, and tangent to the y-axis there, into a unit circle $\tilde{C}$, with Δ_A the arc length on the unit circle of the segment used for normalization (see Fig. 1). Again, the normalizing transformation depends on the choice of the point P in addition to the interval size Δ_A.

To determine the correspondence between target segment and the templates, we now do a search over the templates, over choice of fiducial point P' on each template, and over the two possible senses of the tangent vector at P', applying the same normalization recipe as was applied to the target segment and looking for a match. The search for a match is again done by comparing the normalized target and template using an appropriate measure for the degree to which they coincide.

Variants on this procedure are possible. For instance, we could instead (see reference (1)) use the affine subgroup defined by matrices of the form

$$B = \begin{pmatrix} 1 & 0 \\ \beta & \gamma \end{pmatrix}, \tag{14}$$

which give a general affine transformation when combined with a similarity transformation. In this case one would only require $\mu_{A;02}(T_A) = \mu_{A;20}(T_A)$ as a normalization condition on the moments $\mu_{A;02}$ and $\mu_{A;20}$, without requiring the common value to be unity. One would then have to do a scaling normalization, using the method described in Section 2, following the partial normalization with respect to the affine trans-

formation of equation (14), and this scaling normalization would have to be placed *inside* the Newton iteration loop which determines the integration interval T.

To conclude, we have shown that the normalization methods of reference (1) are not limited to the case of non-occluded images. When applied to partially occluded curves, they give a method for similarity normalization using only the tangent vector (which requires only first parametric derivatives), and a method for affine normalization using only the tangent vector and curvature (which requires only first and second parametric derivatives).

4. SUMMARY

We extend the general theory of image normalization developed by Adler[1] to the case of planar curve segments. Specifically, we show that the method can be used to obtain a normalization, and hence *all* the invariants, of partially occluded planar curves subjected to similarity and affine transformations. The similarity normalization uses only the tangent vector (which requires only first parametric derivatives), while the affine normalization uses only the tangent vector and the curvature (which requires only first and second parametric derivatives). Thus, our algorithm gives a substantial improvement over previous methods in the literature for constructing similarity and affine invariants of partially occluded planar curves. Our algorithm is based on the subgroup method,[1] in which we solve in succession the normalization problem for planar curve segments under translations and rotations, similarity transformations, and finally affine transformations. The similarity normalization is given by an explicit parametric integral over the curve segment; the affine normalization is given in terms of parametric integrals over the curve segment by a single, rapidly convergent Newton iteration. We give all the formulas needed for constructing the normalization algorithm, and a figure illustrating typical computational results.

Acknowledgements—This work was supported in part by the Department of Energy under Grant #DE-FG02-90ER40542. S.L.A. wishes to acknowledge the hospitality of Clare Hall and the Department of Applied Mathematics and Theoretical Physics at Cambridge University, where part of this work was done. He also wishes to thank participants in the Nordfjordeid Machine Vision workshop, and in particular A. M. Bruckstein, for introducing him to the problem discussed here.

APPENDIX A

We derive here a framework for the normalization of planar curves, starting from the general theory of image normalization given in reference (1). Let C be a planar curve segment $\mathrm{x} = \mathrm{x}_C(t)$ parameterized by t, which in the terminology of reference (1) corresponds to an image intensity per unit area $I(\mathrm{x})$ given by the reparameterization invariant expression

$$I(\mathrm{x}) = \int_{t_{\min}}^{t_{\max}} \mathrm{d}t|\mathbf{T}|\delta^2(\mathrm{x} - \mathrm{x}_C(t)). \tag{A1a}$$

In Equation (A1a), $\delta^2(\mathrm{x})$ is the two-dimensional Dirac delta function, defined by $\delta^2(\mathrm{x}) = 0$ for $\mathrm{x} \neq 0$, and $\int\int \mathrm{d}^2\mathrm{x}\, \delta^2(\mathrm{x}) = 1$, so that $\delta^2(\mathrm{x})$ is a distribution of unit weight with support at the origin $\mathrm{x} = 0$. Also in equation (A1a), $\mathbf{T}$ is the tangent vector defined in equation (1b), so that $\mathrm{d}s = \mathrm{d}t|\mathbf{T}|$ is the differential of arc length, and therefore equation (A1a) describes a distribution with support on the curve segment C having uniform weight per unit of arc length. The total image intensity integrated over area, corresponding to equation (A1a), is given by

$$\int\int \mathrm{d}^2\mathrm{x}\, I(\mathrm{x}) = \int_{t_{\min}}^{t_{\max}} \mathrm{d}t|\mathbf{T}| \int\int \mathrm{d}^2\mathrm{x}\, \delta^2(\mathrm{x} - \mathrm{x}_C(t))$$
$$= \int_{t_{\min}}^{t_{\max}} \mathrm{d}t|\mathbf{T}| = s_{\max} - s_{\min}, \tag{A1b}$$

and so is just the total arc length of the segment.

Let now $\mathscr{G} = \{S\}$ be a group of symmetry or viewing transformations S, which act on the image $I(\mathrm{x})$ according to

$$I(\mathrm{x}) \to I_S(\mathrm{x}) = I(\mathbf{S}(\mathrm{x})), \tag{A2a}$$

with $\mathrm{x} \to \mathbf{S}(\mathrm{x})$ the image coordinate mapping induced by the group element S. Substituting Equation (A1a) into equation (A2a), we have

$$I_S(\mathrm{x}) = \int_{t_{\min}}^{t_{\max}} \mathrm{d}t|\mathbf{T}|\delta^2(\mathbf{S}(\mathrm{x}) - \mathrm{x}_C(t))$$
$$= \int_{t_{\min}}^{t_{\max}} \mathrm{d}t|\mathbf{T}||J(\mathrm{x})|^{-1}\delta^2(\mathrm{x} - \mathbf{S}^{-1}(\mathrm{x}_C(t))), \tag{A2b}$$

with $J(\mathrm{x})$ the Jacobian of the transformation $\mathbf{S}(\mathrm{x})$. For the case of similarity and affine transformations discussed in this paper, the Jacobian J is simply a constant and plays no further role; the image transformation of equations (A2a) and (A2b) is then equivalent to the replacement of the curve segment C by the transformed curve segment C_S described by the parametric expression

$$\mathrm{x}_{C_S}(t) = \mathbf{S}^{-1}(\mathrm{x}_C(t)). \tag{A3}$$

Using equation (A3), the various results for image normalization given in reference (1) can be taken over to the planar curve case, with due attention to the fact that the inverse transformation appears in equation (A3). In particular, since

$$\mathbf{f}(\mathbf{g}(\mathrm{x}))^{-1} = \mathbf{g}^{-1}(\mathbf{f}^{-1}(\mathrm{x})), \tag{A4}$$

expressions in reference (1) which involve multiple transformations will have the factors reverse ordered when expressed as an action on the parameterized curve. For example, the general normalization recipe for the parameterized curve reads

$$\tilde{\mathrm{x}}_C(t) = \mathbf{M}_C(\mathbf{N}_C(\mathrm{x}_C(t))), \tag{A5a}$$

where $\mathbf{N}_C(\mathbf{x})$ is a normalizing map constructed from the curve which transforms under the image transformation as

$$\mathbf{N}_{C_S}(\mathbf{x}) = \mathbf{N}_C(\mathbf{S}(\mathbf{x})), \tag{A5b}$$

and where $\mathbf{M}_C(\mathbf{x})$ is an optional second transformation which depends on the curve C only through invariants under the group of transformations $\mathscr{G}$, that is,

$$\mathbf{M}_{C_S}(\mathbf{x}) = \mathbf{M}_C(\mathbf{x}), \quad \text{all } S \in \mathscr{G}. \tag{A5c}$$

We can easily check the validity of equation (A5a) directly,

$$\begin{aligned} \tilde{\mathbf{x}}_{C_S}(t) &= \mathbf{M}_{C_S}(\mathbf{N}_{C_S}(\mathbf{x}_{C_S}(t))) \\ &= \mathbf{M}_C(\mathbf{N}_C(\mathbf{S}(\mathbf{S}^{-1}(\mathbf{x}_C(t))))) \\ &= \mathbf{M}_C(\mathbf{N}_C(\mathbf{x}_C(t))) = \tilde{\mathbf{x}}_C(t). \end{aligned} \tag{A6}$$

The other general statements in reference (1) are similarly converted to results for the normalization of parameterized curves.

APPENDIX B

We give here the central moments for a unit circle passing through the origin and tangent to the y-axis there. The parameterized form for the circle is

$$\mathbf{x}_{\text{unit circle}}(t) = (-1 + \cos t, \sin t), \; 0 \le t < 2\pi, \tag{B1}$$

and the origin lies at $t = 0$. Equation (7) for the affine covariant arc length reduces to $d\tau = dt$, and we choose the constant of integration so that $\tau = 0$ at the origin. Integrating over a segment of the unit circle $0 \le \tau = t \le T$, we find the following formulas for the center of mass coordinates and the second central moments,

$$\begin{aligned} (x_{CM}, y_{CM}) &= T^{-1}(-T + \sin T, 1 - \cos T), \\ \mu_{20}(T) &= \frac{T}{2}\left(1 + \frac{\sin 2T}{2T}\right) - \frac{1}{T}(\sin T)^2, \\ \mu_{02}(T) &= \frac{T}{2}\left(1 - \frac{\sin 2T}{2T}\right) - \frac{4}{T}\left(\sin\frac{T}{2}\right)^4, \\ \mu_{11}(T) &= \sin T \sin\frac{T}{2}\left(\cos\frac{T}{2} - \frac{2}{T}\sin\frac{T}{2}\right), \end{aligned} \tag{B2a}$$

and the corresponding small T behavior is

$$\begin{aligned} (x_{CM}, y_{CM}) &\approx \left(-\frac{1}{6}T^2, \frac{1}{2}T\right), \\ \mu_{20}(T) &\approx \frac{1}{45}T^5 - \frac{1}{315}T^7, \\ \mu_{02}(T) &\approx \frac{1}{12}T^3 - \frac{1}{40}T^5, \\ \mu_{11}(T) &\approx -\frac{1}{24}T^4 + \frac{7}{720}T^6, \\ \mathscr{D}^{-1/8}(T) &= [\mu_{20}(T)\mu_{02}(T) - \mu_{11}(T)^2]^{-1/8} \\ &\approx 8640^{1/8}T^{-1}\left(1 + \frac{3}{280}T^2\right). \end{aligned} \tag{B2b}$$

REFERENCES

1. S. L. Adler, General theory of image normalization, *Computer Vision and Image Understanding*, submitted. (Also a Lecture at Int. Workshop on Computer Vision and Applied Geometry, Nordfjordeid, Norway, 1–7 August, 1995.)
2. A. M. Bruckstein and A. N. Netravali, On differential invariants of planar curves and recognizing partially occluded planar shapes, *Ann. Math. Artificial Intell.* **13**, 227–250 (1995).
3. A. M. Bruckstein, E. Rivlin and I. Weiss, Scale space semilocal invariants. *Image Vision Comput.* **15**, 335–344 (1997).
4. R. F. Vaz and D. Cyganski, *Pattern Recognition Letters* **11**, 479–483 (1990).

About the Author—STEPHEN L. ADLER received his A.B. from Harvard University in 1961, his Ph.D. in theoretical physics from Princeton University in 1964, and was a Junior Fellow in the Society of Fellows, Harvard University, from 1964 to 1966. Since 1966 he has been at the Institute for Advanced Study in Princeton, as a Long Term Member (1966–69), Professor (1969–79), and New Jersey Albert Einstein Professor (1979–). His research interests include neural networks and pattern recognition, as well as continued work in theoretical high energy physics.

About the Author—RANGANATHAN KRISHNAN received his B.S. from the Indian Institute of Technology in Madras in 1989, his Ph.D. in theoretical physics from the Massachusetts Institute of Technology in 1994, and was a Member of the Institute for Advanced Study in Princeton from 1964 to 1966. He is currently employed at Dragon Systems, in Newton MA, working on computer methods for speech recognition.

Reprinted from FOUNDATIONS OF PHYSICS Vol. 26, No. 12, December 1996
Printed in Belgium

Nonadiabatic Geometric Phase in Quaternionic Hilbert Space

Stephen L. Adler[1] **and Jeeva Anandan**[2]

Received September 30, 1996; revised October 25, 1996

We develop the theory of the nonadiabatic geometric phase, in both the Abelian and non-Abelian cases, in quaternionic Hilbert space.

1. INTRODUCTION

The theory of geometric phases associated with cyclic evolutions of a physical system is now a well-developed subject in complex Hilbert space. The seminal work of Berry on the adiabatic single state (Abelian) case[(1)] has been extended to the non-Abelian case of the adiabatic evolution of a set of degenerate states,[(2)] and both of these have been further extended[(3, 4)] to show that there is a geometric phase associated with any cyclic but non-adiabatic evolution of a single quantum state or of a degenerate group of quantum states.

In this paper we take up another direction for generalization of the geometric phase, from quantum mechanics in complex Hilbert space to quantum mechanics[(5, 6)] in quaternionic Hilbert space. The generalization of the adiabatic geometric phase to quaternionic Hilbert space was given in Ref. 6, where it was shown that for states of nonzero energy the adiabatic geometric phase is complex, as opposed to quaternionic, with a quaternionic adiabatic geometric phase occurring only for the adiabatic cyclic

[1] Institute for Advanced Study, Princeton, New Jersey 08540.
[2] Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208.

evolution of zero energy states. Consideration of nonadiabatic cyclic evolutions was also begun in Ref. 6, but the discussion given there is incomplete. While Sec. 5.8 of Ref. 6 constructed a nonadiabatic cyclic invariant phase, it did not address the problem of separating this phase into a dynamical part determined by the quantum mechanical Hamiltonian, and a geometric part that depends only on the ray orbit and is independent of the Hamiltonian.

The purpose of the present paper is to give a complete discussion of the nonadiabatic geometric phase in quaternionic Hilbert space. In Sec. 2 we give a very brief survey of the properties of quantum mechanics in quaternionic Hilbert space that are needed in the analysis that follows. In Sec. 3 we consider the cyclic nonadiabatic evolution of a single quantum state, and show how to explictly generalize to quaternionic Hilbert space the construction of a nonadiabatic geometric phase given in Ref. 3. In Sec. 4 we extend our analysis to the case of a degenerate group of states, thereby obtaining a quaternionic nonadiabatic non-Abelian geometric phase corresponding to the complex construction given in Ref. 4. A brief summary and discussion of our results is given in Sec. 5.

2. QUANTUM MECHANICS IN QUATERNIONIC HILBERT SPACE

Only a few properties of quaternionic quantum mechanics are needed for the discussion that follows; the reader wishing to learn more than we can present here should consult Ref. 6. In quaternionic quantum mechanics, the Dirac transition amplitudes $\langle\psi|\phi\rangle$ are quaternion valued, that is, they have the form

$$\langle\psi|\phi\rangle = r_0 + r_1 i + r_2 j + r_3 k \tag{1}$$

where $r_{0,1,2,3}$ are real numbers and where i, j, k are quaternion imaginary units obeying the associative algebra $i^2 = j^2 = k^2 = -1$ and $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. Because quaternion multiplication is noncommutative, two independent Dirac transition amplitudes $\langle\psi|\phi\rangle$ and $\langle\kappa|\eta\rangle$ in general do not commute with one another, unlike the situation in standard complex quantum mechanics, where all Dirac transition amplitudes are complex numbers and mutually commute. The transition probability corresponding to the amplitude of Eq. (1) is given by

$$P(\psi, \phi) = |\langle\psi|\phi\rangle|^2 = \overline{\langle\psi|\phi\rangle}\,\langle\psi|\phi\rangle = r_0^2 + r_1^2 + r_2^2 + r_3^2 \tag{2}$$

where the bar denotes the quaternion conjugation operation $\{i, j, k\} \to \{-i, -j, -k\}$ and where we have assumed the states $|\psi\rangle$ and $|\phi\rangle$ to be

unit normalized. Since the quaternion norm defined by Eq. (2) has the multiplicative norm property

$$|q_1 q_2| = |q_1|\,|q_2| \tag{3}$$

the transition probability of Eq. (2) is unchanged when the state vector $|\phi\rangle$ is right multiplied by a quaternion ω of unit magnitude,

$$|\phi\rangle \to |\phi\rangle\,\omega, \qquad |\omega| = 1 \Rightarrow P(\psi, \phi) \to P(\psi, \phi) \tag{4}$$

Hence as in complex quantum mechanics, physical states are associated with Hilbert space rays of the form $\{|\phi\rangle\,\omega\colon |\omega| = 1\}$, and the transition probability of Eq. (2) is the same for any ray representative state vectors $|\phi\rangle$ and $|\psi\rangle$ chosen from their corresponding rays. In the next section, we shall follow Ref. 3 in denoting quaternionic Hilbert space by $\mathscr{H}$, and the projective Hilbert space of rays of $\mathscr{H}$ by $\mathscr{P}$.

Time evolution of the state vector $|\psi\rangle$ is described in quaternionic quantum mechanics by the Schrödinger equation

$$\frac{\partial\,|\psi\rangle}{\partial t} = -\tilde{H}\,|\psi\rangle \tag{5a}$$

with

$$\tilde{H} = -\tilde{H}^{\dagger} \tag{5b}$$

an anti-self-adjoint Hamiltonian. From Eqs. (5a) and (5b) we see that the Dirac transition amplitude $\langle\psi|\phi\rangle$ is time independent,

$$\begin{aligned}\frac{\partial}{\partial t}\langle\psi|\phi\rangle &= \left(\frac{\partial}{\partial t}\langle\psi|\right)|\phi\rangle + \langle\psi|\,\frac{\partial}{\partial t}\,|\phi\rangle \\ &= \langle\psi|\,\tilde{H} - \tilde{H}\,|\phi\rangle = 0 \end{aligned}\tag{6}$$

and thus the Schrödinger dynamics of state vectors preserves the inner product structure of Hilbert space. The dynamics of Eqs. (5) and (6) is evidently preserved under right linear superposition of states with quaternionic constants,

$$\begin{aligned}&\frac{\partial\,|\psi\rangle}{\partial t} = -\tilde{H}\,|\psi\rangle, \quad \frac{\partial\,|\phi\rangle}{\partial t} = -\tilde{H}\,|\phi\rangle \\ &\Rightarrow \frac{\partial(|\psi\rangle\,q_1 + |\phi\rangle\,q_2)}{\partial t} \\ &= -\tilde{H}(|\psi\rangle\,q_1 + |\phi\rangle\,q_2)\end{aligned}\tag{7}$$

Equation (7) illustrates two general features of our conventions for quaternionic quantum mechanics, which are that linear operators (such as $\tilde{H}$) act on Hilbert space state vectors by multiplication from the left, whereas quaternionic numbers (the scalars of Hilbert space) act on state vectors by multiplication from the right. Adherence to these ordering conventions is essential because of the noncommutative nature of quaternionic multiplication.

3. THE NONADIABATIC ABELIAN QUATERNIONIC GEOMETRIC PHASE

Let us now consider a unit normalized quaternionic Hilbert space state $|\psi(t)\rangle$ which undergoes a cyclic evolution between the times $t=0$ and $t=T$. Since physical states are associated with rays, this means that

$$|\psi(T)\rangle = |\psi(0)\rangle\, \Omega, \qquad |\Omega| = 1 \tag{8}$$

and so the orbit $\mathscr{C}$ of $|\psi(t)\rangle$ in $\mathscr{H}$ projects to a closed curve $\hat{\mathscr{C}}$ in the projective Hilbert space $\mathscr{P}$.

Let us now define a state $|\hat{\psi}(t)\rangle$ that is equal to $|\psi(t)\rangle$ at $t=0$, that differs from $|\psi(t)\rangle$ only by a reraying at general times, i.e.,

$$\begin{aligned} &|\psi(t)\rangle = |\hat{\psi}(t)\rangle\, \hat{\omega}(t) \\ &|\hat{\omega}(t)| = 1 \\ &\hat{\omega}(0) = 1 \end{aligned} \tag{9a}$$

and that evolves in time by parallel transport, i.e.,

$$\langle \hat{\psi}(t)|\, \frac{\partial\, |\hat{\psi}(t)\rangle}{\partial t} = 0 \tag{9b}$$

The conditions of Eqs. (9a) and (9b) uniquely determine $\hat{\omega}(t)$, and hence the state $|\hat{\psi}(t)\rangle$, as follows. Substituting the first line of Eq. (9a) into the Schrödinger equation of Eq. (5a), we get

$$\begin{aligned} -\tilde{H}\, |\hat{\psi}(t)\rangle\, \hat{\omega}(t) &= -\tilde{H}\, |\psi(t)\rangle \\ &= \frac{\partial\, |\psi(t)\rangle}{\partial t} = |\hat{\psi}(t)\rangle\, \frac{d\hat{\omega}(t)}{dt} + \frac{\partial\, |\hat{\psi}(t)\rangle}{\partial t}\, \hat{\omega}(t) \end{aligned} \tag{10}$$

Taking the inner product of this equation with the state $\langle\hat{\psi}(t)|$, and using the unit normalization of the state vector $|\hat{\psi}(t)\rangle$ together with the parallel transport condition of Eq. (9b), we get

$$\frac{d\hat{\omega}(t)}{dt} = -\langle\hat{\psi}(t)|\, \tilde{H}\, |\hat{\psi}(t)\rangle\, \hat{\omega}(t) \tag{11}$$

This differential equation can be immediately integrated to give

$$\hat{\omega}(t) = T_t e^{-\int_0^t dv\, \langle\hat{\psi}(v)|\, \tilde{H}\, |\hat{\psi}(v)\rangle} \tag{12}$$

where T_t denotes the time-ordered product which orders later times to the left, and where we have used the initial condition on the third line of Eq. (9a). In particular, Eq. (12) gives us a formula for the value $\hat{\omega}(T)$ at the end of the cyclic evolution. We shall see that this has the interpretation of the dynamics-dependent part of the total phase change Ω.

To relate Eq. (12) to the total phase change, we use Eqs. (8) and (9a) to write

$$|\hat{\psi}(T)\rangle\, \hat{\omega}(T) = |\psi(T)\rangle = |\psi(0)\rangle\, \Omega = |\hat{\psi}(0)\rangle\, \Omega \tag{13a}$$

so that taking the inner product with $\langle\hat{\psi}(0)|$ gives

$$\Omega = \langle\hat{\psi}(0)|\, \hat{\psi}(T)\rangle\, \hat{\omega}(T) \tag{13b}$$

To complete the calculation, we must now evaluate the inner product appearing in Eq. (13b). To do this, we introduce a third state vector $|\tilde{\psi}(t)\rangle$ which differs from $|\hat{\psi}(t)\rangle$ by a change of ray representative, by writing

$$\begin{aligned} &|\hat{\psi}(t)\rangle = |\tilde{\psi}(t)\rangle\, \tilde{\omega}(t) \\ &|\tilde{\omega}(t)| = 1 \\ &\tilde{\omega}(0) = 1 \end{aligned} \tag{14a}$$

and by requiring that $\tilde{\psi}$ should be continuous over the orbit $\mathscr{C}$,

$$|\tilde{\psi}(T)\rangle = |\tilde{\psi}(0)\rangle \tag{14b}$$

Differentiating the first line of Eq. (14a) with respect to time, we get

$$\frac{\partial\, |\hat{\psi}(t)\rangle}{\partial t} = \frac{\partial\, |\tilde{\psi}(t)\rangle}{\partial t}\, \tilde{\omega}(t) + |\tilde{\psi}(t)\rangle\, \frac{d\tilde{\omega}(t)}{dt} \tag{15}$$

Taking the inner product of Eq. (15) with $\tilde{\omega}(t)\,\langle\hat{\psi}(t)|$, using the parallel transport condition of Eq. (9b) together with the first line of Eq. (14a), and abbreviating the time derivative $\partial/\partial t$ by a dot, we obtain

$$0=\langle\tilde{\psi}(t)|\,\dot{\tilde{\psi}}\,|(t)\rangle\,\tilde{\omega}(t)+\langle\tilde{\psi}(t)\,|\,\tilde{\psi}(t)\rangle\,\dot{\tilde{\omega}}(t)) \tag{16a}$$

Since the second line of Eq. (14a) implies that the state $|\hat{\psi}(t)\rangle$ is unit normalized, Eq. (16a) simplifies to

$$\dot{\tilde{\omega}}(t))=-\langle\tilde{\psi}(t)|\,\dot{\tilde{\psi}}(t)\rangle\,\tilde{\omega}(t) \tag{16b}$$

which can be immediately integrated to give

$$\tilde{\omega}(t)=T_t e^{-\int_0^t dv\langle\tilde{\psi}(v)|\,\dot{\tilde{\psi}}(v)\rangle} \tag{17}$$

with T_t as before indicating a time-ordered product. In particular, Eq. (17) gives us a formula for $\tilde{\omega}(T)$. But from Eqs. (14a) and (14b) we have

$$|\hat{\psi}(T)\rangle=|\tilde{\psi}(T)\rangle\,\tilde{\omega}(T)=|\tilde{\psi}(0)\rangle\,\tilde{\omega}(T)=|\hat{\psi}(0)\rangle\,\tilde{\omega}(T) \tag{18a}$$

and so taking the inner product of Eq. (18a) with $\langle\hat{\psi}(0)|$ we get

$$\langle\hat{\psi}(0)|\,\hat{\psi}(T)\rangle=\tilde{\omega}(T) \tag{18b}$$

determining the inner product appearing in Eq. (13b).

We thus get as our final result,

$$\Omega=\Omega_{\text{geometric}}\Omega_{\text{dynamical}} \tag{19a}$$

with

$$\Omega_{\text{geometric}}\equiv\tilde{\omega}(T)=T_t e^{-\int_0^T dv\,\langle\tilde{\psi}(v)|\,\dot{\tilde{\psi}}(v)\rangle} \tag{19b}$$

and with

$$\Omega_{\text{dynamical}}\equiv\hat{\omega}(T)=T_t e^{-\int_0^T dv\,\langle\hat{\psi}(v)|\,\tilde{H}\,|\hat{\psi}(v)\rangle} \tag{19c}$$

The dynamical part of the phase is so called because it depends explicitly on $\tilde{H}$, as well as on the orbit $\mathscr{C}$ in the projective Hilbert space $\mathscr{P}$; it is uniquely determined by the conditions of Eqs. (9a) and (9b), since these conditions uniquely determine the state $|\hat{\psi}(t)\rangle$. The geometric part of the phase is so called because, as we shall now show, it depends uniquely on

the projective orbit $\hat{\mathscr{C}}$ up to an overall quaternion automorphism transformation. To see this, let us make the reraying

$$|\tilde{\psi}(t)\rangle \to |\tilde{\psi}'\rangle\, \omega'(t), \qquad |\omega'| = 1 \tag{20a}$$

with $\omega'(t)$ continuous over the orbit $\mathscr{C}$ so that

$$\omega'(T) = \omega'(0) \tag{20b}$$

Then (as shown in detail in Sec. 5.8 of Ref. 6) the properties of the time-ordered integral in Eq. (19b) imply that under this transformation,

$$\Omega_{\text{geometric}} \to \bar{\omega}'(T)\, \Omega_{\text{geometric}}\, \omega'(0) \tag{21a}$$

which by the continuity condition of Eq. (20b) reduces to the quaternion automorphism transformation

$$\Omega_{\text{geometric}} \to \bar{\omega}'(0)\, \Omega_{\text{geometric}}\, \omega'(0) \tag{21b}$$

Since for any two quaternions q_1, q_2 we have $\text{Re}\, q_1 q_2 = \text{Re}\, q_2 q_1$, with Re denoting the real part, Eq. (21b) implies that

$$\cos \gamma_{\text{geometric}} \equiv \text{Re}\, \Omega_{\text{geometric}} \tag{22}$$

is a reraying invariant, and thus $\gamma_{\text{geometric}}$ is a nonadiabatic geometric phase angle that is a property solely of the projective orbit $\hat{\mathscr{C}}$. The fact that the nonadiabatic geometric phase in quaternionic Hilbert space is only determined modulo π is a reflection of the fact that $e^{i\gamma}$ is changed to $e^{-i\gamma}$ by the quaternion automorphism transformation

$$e^{-i\gamma} = \bar{j} e^{i\gamma} j \tag{23}$$

Thus, to recover the result that the complex nonadiabatic geometric phase is determined modulo 2π by embedding a complex Hilbert space in a quaternionic one and using Eqs. (19a)–(19c), one must exclude the possibility of making intrinsically quaternionic automorphism transformations involving the quaternion units j or k, as in Eq. (23).

In geometric terms, $\Omega_{\text{geometric}}$ is the holonomy transformation of the connection $A \equiv \langle \tilde{\psi} | d\tilde{\psi} \rangle$. But since this connection is quaternion-imaginary valued, it is analogous to an $SO(3)$ gauge potential. Therefore, the corresponding curvature is of the Yang–Mills type and is given by $F = dA + A \wedge A$.

An alternative expression for the total phase change Ω can be obtained[7] by writing

$$\begin{aligned} |\psi(t)\rangle &= |\tilde{\psi}(t)\rangle\, \tilde{\chi}(t) \\ \tilde{\chi}(t) &= \hat{\omega}(t)\, \tilde{\omega}(t), \qquad \tilde{\chi}(0) = 1 \end{aligned} \tag{24}$$

Substituting Eq. (24) into the Schrödinger equation and then taking the inner product with $\langle \tilde{\psi}(t)|$, we obtain

$$\frac{d\tilde{\chi}(t)}{dt} = -(\langle \tilde{\psi}(t)|\, \tilde{H}\, |\tilde{\psi}(t)\rangle + \langle \tilde{\psi}(t)|\, \dot{\tilde{\psi}}(t)\rangle)\, \tilde{\chi}(t) \tag{25a}$$

which can be integrated from 0 to T to give

$$\Omega = T_{l} e^{-\int_0^T dv (\langle \tilde{\psi}(v)|\, \tilde{H}\, |\tilde{\psi}(v)\rangle + \langle \tilde{\psi}(v)|\, \dot{\tilde{\psi}}(v)\rangle)} \tag{25b}$$

This procedure and the resulting formula of Eq. (25b) are direct analogs of the derivation given in Ref. 3 for the complex Hilbert space case, but in quaternionic Hilbert space the two terms in the exponential are noncommutative, and so the exponential in Eq. (25b) cannot be immediately factored into dynamical and geometric phase factors. As we have seen, to achieve this factorization it is necessary to use a two-step procedure, involving the parallel transported state $|\hat{\psi}(t)\rangle$ as well as the state $|\tilde{\psi}(t)\rangle$ that is continuous over the cycle.

4. THE NONADIABATIC NON-ABELIAN QUATERNIONIC GEOMETRIC PHASE

We turn next to the quaternionic Hilbert space generalization of the complex nonadiabatic[4] non-Abelian[2] geometric phase. We consider now a cyclic evolution in an n-dimensional Hilbert subspace V_n, i.e., $V_n(T) = V_n(0)$. Let $|\psi_a(t)\rangle$, $a = 1, \ldots, n$ be a complete orthonormal basis for V_n, so that the reraying invariant projection operator for V_n is

$$\rho_n(t) = \sum_{a=1}^{n} |\psi_a(t)\rangle\langle \psi_a(t)| \tag{26a}$$

in terms of which the cyclic evolution condition takes the form

$$\rho_n(T) = \rho_n(0) \tag{26b}$$

Expressed in terms of the state vectors of the basis, the invariance of V_n implies that the basis element $|\psi_b(T)\rangle$ must be a superposition of the basis elements $|\psi_a(0)\rangle$, multiplied from the right by quaternionic coefficients U_{ab},

$$|\psi_b(T)\rangle = \sum_{a=1}^{n} |\psi_a(0)\rangle\, U_{ab} \tag{27}$$

Substituting Eq. (27) into Eqs. (26a, b), we find that invariance of the projection operator requires

$$\begin{aligned} \rho_n(T) &= \sum_{b=1}^{n} |\psi_b(T)\rangle\langle\psi_b(T)| \\ &= \sum_{a,b,c=1}^{n} |\psi_a(0)\rangle\, U_{ab}\bar{U}_{cb}\,\langle\psi_c(0)| \\ &= \sum_{a=1}^{n} |\psi_a(0)\rangle\langle\psi_a(0)| = \rho_n(0) \end{aligned} \tag{28a}$$

which implies that

$$\sum_{b=1}^{n} U_{ab}\bar{U}_{cb} = \delta_{ac} \tag{28b}$$

Similarly, orthonormality of the basis at $t=0$ and $t=T$ implies that

$$\begin{aligned} \delta_{ab} = \langle\psi_a(T)\mid\psi_b(T)\rangle &= \sum_{c,d=1}^{n} \bar{U}_{ca}\,\langle\psi_c(0)\mid\psi_d(0)\rangle\, U_{db} \\ &= \sum_{c,d=1}^{n} \bar{U}_{ca}\,\delta_{cd}\,U_{db} = \sum_{d=1}^{n} \bar{U}_{da}\,U_{db} \end{aligned} \tag{28c}$$

Hence U is an $n\times n$ quaternion unitary matrix, $U^\dagger U = UU^\dagger = 1$, which replaces the quaternion phase Ω (which is a 1×1 quaternion unitary matrix) of the preceding section. Thus, to generalize the results of the preceding section to the non-Abelian case (i) one replaces the state vectors $|\psi\rangle$, $|\hat{\psi}\rangle$, $|\tilde{\psi}\rangle$ by n-component column vectors $|\psi_a\rangle$, $|\hat{\psi}_a\rangle$, $|\tilde{\psi}_a\rangle$, $a=1,\dots,n$, (ii) one replaces the phases $\Omega, \hat{\omega},\dots$ by $n\times n$ quaternion unitary matrices acting on the state vector indices, (iii) one replaces Re in Eq. (22) by Re Tr, with Tr the trace over the subspace V_n, and (iv) as in Ref. 4, one generalizes the parallel transport condition of Eq. (9b) to

$$\langle\hat{\psi}_a(t)|\,\frac{\partial\,|\hat{\psi}_b(t)\rangle}{\partial t} = 0, \qquad a, b = 1,\dots, n \tag{28d}$$

The principal difference from the complex case treated in Ref. 4 is that in the quaternion case, the unitary matrix factors must always be ordered to the right of ket state vectors, whereas in the complex case the ordering is irrelevant, and in fact in Ref. 4 the matrix factors are ordered to the left. The results of Ref. 4 can be obtained by the complex specialization of the results obtained in this paper. However, we have introduced here a new technique of using parallel transported states $|\hat{\psi}_a\rangle$ to cleanly separate the non-Abelian geometric phase and the dynamical phase, which in general (even in the complex non-Abelian case) do not commute with each other.

5. SUMMERY AND DISCUSSION

To summarize, we have shown that both the complex Abelian and non-Abelian nonadiabatic geometric phases can be generalized to quaternonic Hilbert space. These results are both of theoretical interest and of experimental relevance for possible tests for complex versus quaternionic quantum mechanics. Long ago, Peres[8] proposed testing for quaternionic quantum mechanical effects by looking for noncommutativity of scattering phase shifts. However, the result of Ref. 6 that the S-matrix in quaternionic quantum mechanics is always complex valued (for nonzero energy states) implies that there are no quaternionic scattering phase shifts, and the Peres test necessarily gives a null result. An alternative but related method is to look for interference effects in cyclic evolutions that could show the presence of quaternionic effects. The fact[6] that the adiabatic geometric phase is always complex (for nonzero energy states) is a counterpart of the complexity of the S-matrix, and implies that a null result will always be obtained for cyclic interference experiments involving adiabatic state evolutions. However, the results obtained here show that for cyclic evolutions that are nonadiabatic, one could in principle devise interference experiments to place meaningful bounds on postulated quaternionic components of the wave function.

ACKNOWLEDGMENTS

The work of SLA was supported in part by the Department of Energy under Grant No. DE-FG02-90ER40542, and of JA in part by ONR grant No. R&T 3124141 and NSF grant No. PHY-9307708. SLA wishes to acknowledge the hospitality of the Aspen Center for Physics, where part of this work was done.

REFERENCES

1. M. V. Berry, *Proc. R. Soc. London A* **392**, 45 (1984).
2. F. Wilczek and A. Zee, *Phys. Rev. Lett.* **56**, 2111 (1984).
3. Y. Aharanov and J. Anandan, *Phys. Rev. Lett.* **58**, 1593 (1987).
4. J. Anandan, *Phys. Lett. A* **133**, 171 (1988).
5. L. P. Horwitz and L. C. Biedenharn, *Ann. Phys. New York* **157**, 432 (1984).
6. S. L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford University Press, New York and Oxford, 1995).
7. P. Lévay, *Phys. Rev. A* **41**, 2837 (1990); *J. Math. Phys.* **32**, 2347 (1991).
8. A. Peres, *Phys. Rev. Lett.* **42**, 683 (1979).

Coherent states in quaternionic quantum mechanics

Stephen L. Adler[a)]
Institute for Advanced Study, Princeton, New Jersey 08540
Andrew C. Millard[b)]
Physics Department, Princeton University, Princeton, New Jersey 08544

(Received 28 October 1996; accepted for publication 2 December 1996)

We develop Perelomov's coherent states formalism to include the case of a quaternionic Hilbert space. We find that, because of the closure requirement, an attempted quaternionic generalization of the special nilpotent or Weyl group reduces to the normal complex case. For the case of the compact group SU(2), however, coherent states can be formulated using the quaternionic half-integer spin matrices of Finkelstein, Jauch, and Speiser, giving a nontrivial quaternionic analog of coherent states. © *1997 American Institute of Physics.* [S0022-2488(97)01005-0]

I. INTRODUCTION

The coherent states formalism is an important part of the apparatus of complex quantum mechanics, and in this framework has been given a general and elegant form through the work of Perelomov.[1] However, in a recent systematic study of quantum mechanics in quaternionic Hilbert space,[2] the issue of whether there is a quaternionic analog of coherent states was left open; filling this gap is the object of the present paper. In Sec. II we show that the general Perelomov construction readily extends to quaternionic Hilbert space, even when the subtleties arising from projective group representations[3] are taken into account. In Sec. III we demonstrate that when this quaternionic generalization is applied to the special nilpotent or Weyl group, the requirement of group closure reduces the structure of the coherent states so obtained to a quaternionic embedding of the standard complex construction. Hence, as suspected by Klauder,[4] there is no nontrivial quaternionic generalization of the standard complex coherent states based on the Weyl group. As an application of our formalism to a case in which the quaternionic coherent states are not simply embeddings of the corresponding complex ones, we discuss in Sec. IV the case of the quaternionic coherent states constructed by the Perelomov method based on the intrinsically quaternionic half-integer spin representations of the rotation group.

II. GENERAL PROPERTIES OF PERELOMOV COHERENT STATES IN QUATERNIONIC HILBERT SPACE

Let $|\psi_0\rangle$ be a fixed state in a quaternionic Hilbert space $V_{\mathcal{H}}$. For some Lie group G and its irreducible unitary representation, $\{T(g):g\in G\}$, consider the set of states $\{|\psi_g\rangle\}$, where

$$|\psi_g\rangle = T(g)|\psi_0\rangle.$$

Consider transforming from a state $|\psi_{g_1}\rangle$ to another $|\psi_{g_2}\rangle$. In terms of $|\psi_{g_1}\rangle$,

$$|\psi_0\rangle = T^{-1}(g_1)|\psi_{g_1}\rangle = T(g_1^{-1})|\psi_{g_1}\rangle,$$

hence,

[a)]Electronic mail address: adler@ias.edu
[b)]Electronic mail address: amillard@phoenix.princeton.edu

$$|\psi_{g_2}\rangle = T(g_2)T(g_1^{-1})|\psi_{g_1}\rangle = T(g_2 g_1^{-1})|\psi_{g_1}\rangle \omega_p(g_2,g_1^{-1}),$$

where the factor of $\omega_p(g_2,g_1^{-1})$ is a quaternionic phase, inserted so that projective representations[3] may also be considered in this approach. Then, if

$$T(g_2 g_1^{-1})|\psi_{g_1}\rangle = |\psi_{g_1}\rangle \omega_{g_1}(g_2,g_1^{-1}),$$

where $\omega_{g_1}(g_2,g_1^{-1})$ is another quaternionic phase, then

$$|\psi_{g_2}\rangle = |\psi_{g_1}\rangle \omega_{g_1}(g_2,g_1^{-1})\omega_p(g_2,g_1^{-1}),$$

or, in other words, $|\psi_{g_2}\rangle$ and $|\psi_{g_1}\rangle$ differ only by a phase factor and hence determine the same physical state.

Let H be the set of elements $\{h\}$ in G such that

$$T(h)|\psi_0\rangle = |\psi_0\rangle \omega(h).$$

Then H is a subgroup of G, being the stationary group for the ray containing $|\psi_0\rangle$. Forming the set of left cosets $M = G/H$, for each coset $x \in M$, one representative $g(x)$ can be selected to form the set of states $\{|\psi_{g(x)}\rangle\} = \{|x\rangle\}$. The following definition may then be made:

Definition 1: The system of coherent states of type $(T,|\psi_0\rangle)$ is the set of states $\{|\psi_g\rangle\}$, where $|\psi_g\rangle = T(g)|\psi_0\rangle$ and g runs over G. The coherent state $|\psi_g\rangle$ is determined up to a quaternionic phase by the coset $x = x(g)$, which is an element of G/H, corresponding to the element g; that is

$$|\psi_g\rangle = |x> \omega(g),$$

where $|\psi_0\rangle$ is henceforth abbreviated as $|0\rangle$.

Consider h_1, $h_2 \in H$. A general element of G is $g = g(x)h$, where $g(x)$ is a particular element corresponding to a coset in G/H and $g(0) = 1$. From before,

$$|\psi_g\rangle = T(g)|0\rangle = |x\rangle \omega(g),$$

so for $g_1 = g(x)h_1$ and $g_2 = g(x)h_2$,

$$T(g_1)|0\rangle = |x\rangle \omega(x,h_1) \tag{1}$$

and

$$T(g_2)|0\rangle = |x\rangle \omega(x,h_2); \tag{2}$$

similarly, if

$$g_{12} = g(x)h_1 h_2 = g_1 h_2,$$

then

$$T(g_{12})|0\rangle = |x\rangle \omega(x,h_1 h_2). \tag{3}$$

Now consider the case where $|x\rangle = |0\rangle$; Eqs. (1)–(3) then become

$$T(g_1)|0\rangle = |0\rangle \omega(h_1), \quad T(g_2)|0\rangle = |0\rangle \omega(h_2),$$

$$T(g_{12})|0\rangle = |0\rangle \omega(h_1 h_2).$$

However, since $g(0)=1$ implies that $g_1=h_1$ and $g_2=h_2$, then $g_{12}=g_1h_2=g_1g_2$; allowing for projective representations,

$$T(g_{12})|0\rangle=T(g_1g_2)|0\rangle=T(g_1)T(g_2)|0\rangle\omega_p^{-1}(g_1,g_2)=T(g_1)T(g_2)|0\rangle\omega_p^{-1}(h_1,h_2)$$
$$=T(g_1)|0\rangle\omega(h_2)\omega_p^{-1}(h_1,h_2)=|0\rangle\omega(h_1)\omega(h_2)\omega_p^{-1}(h_1,h_2),$$

giving

$$\omega(h_1h_2)=\omega(h_1)\omega(h_2)\omega_p^{-1}(h_1,h_2). \tag{4}$$

If T is a true representation, as opposed to a projective representation, then the projective phases are unity and

$$\omega(h_1h_2)=\omega(h_1)\omega(h_2),$$

in correspondence with the complex phase relationship

$$\exp[i\alpha(h_1h_2)]=\exp[i\alpha(h_1)]\exp[i\alpha(h_2)]$$

given by Perelomov.

Consider now elements $g\in G$ and $h\in H$ and the action of the corresponding operators on $|0\rangle$. Now

$$T(h)|0\rangle=|0\rangle\omega(h)$$

and

$$T(g)|0\rangle=|x(g)\rangle\omega(g),$$

and similarly

$$T(gh)|0\rangle=|x(gh)\rangle\omega(gh);$$

then,

$$T(gh)|0\rangle=T(g)T(h)|0\rangle\omega_p^{-1}(g,h)=|x(g)\rangle\omega(g)\omega(h)\omega_p^{-1}(g,h),$$

and since $x(gh)=x(g)$, this means that

$$\omega(gh)=\omega(g)\omega(h)\omega_p^{-1}(g,h), \tag{5}$$

which is the same phase relationship as in Eq. (4) but with one of the elements of G now not in H.

Finally consider the action of an arbitrary operator, $T(g')$, on an arbitrary coherent state, $|x\rangle$. This may be written

$$T(g')|x\rangle=T(g')|x(g)\rangle=T(g')T(g)|0\rangle\omega^{-1}(g)=T(g'g)|0\rangle\omega_p(g',g)\omega^{-1}(g)$$
$$=|x(g'g)\rangle\omega(g'g)\omega_p(g',g)\omega^{-1}(g)=|x(g'g)\rangle\theta(g',g), \tag{6}$$

where we have defined the new phase

$$\theta(g',g)=\omega(g'g)\omega_p(g',g)\omega^{-1}(g).$$

Replacing g by gh, where h is an element of H, and using Eq. (5) gives

$$\theta(g',gh)=\omega(g'gh)\omega_p(g',gh)\omega^{-1}(gh)$$
$$=\omega(g'g)\omega(h)\omega_p^{-1}(g'g,h)\omega_p(g',gh)\omega_p(g,h)\omega^{-1}(h)\omega^{-1}(g);$$

from the associativity condition for projective representations,

$$\omega_p(g',gh)\omega_p(g,h)=\omega_p(g'g,h)\omega^{-1}(h)\omega_p(g',g)\omega(h),$$

we see that the middle five factors of the last expression for $\theta(g',gh)$ are simply equal to $\omega_p(g',g)$, giving

$$\theta(g',gh)=\omega(g'g)\omega_p(g',g)\omega^{-1}(g).$$

Hence, changing g to gh, where h is any element of H, gives the same θ, so it may be written $\theta(g',x)$, since it only depends on the coset $x(g)$ and not on g itself.

Writing two coherent states as

$$|x_1\rangle=|x(g_1)\rangle=T(g_1)|0\rangle\omega^{-1}(g_1)=T(g_1)|0\rangle\bar{\omega}(g_1),$$

$$|x_2\rangle=|x(g_2)\rangle=T(g_2)|0\rangle\omega^{-1}(g_2)=T(g_2)|0\rangle\bar{\omega}(g_2),$$

their inner product is

$$\langle x_1|x_2\rangle=\langle x(g_1)|x(g_2)\rangle$$
$$=\omega(g_1)\langle 0|T(g_1^{-1})T(g_2)|0\rangle\bar{\omega}(g_2)=\omega(g_1)\langle 0|T(g_1^{-1}g_2)|0\rangle\omega_p(g_1^{-1},g_2)\bar{\omega}(g_2)$$
$$=\omega(g_1)\bar{\omega}_p(g_2^{-1},g_1)\langle 0|T(g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2);$$

replacing g_1 by g_1h, where h is an element of H, in the last line gives

$$\langle x_1|x_2\rangle=\omega(g_1h)\bar{\omega}_p(g_2^{-1},g_1h)\langle 0|T(h^{-1}g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2)$$
$$=\omega(g_1)\omega(h)\omega_p^{-1}(g_1,h)\omega_p^{-1}(g_2^{-1},g_1h)\omega_p(g_2^{-1}g_1,h)\langle 0|T(h^{-1})T(g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2)$$
$$=\omega(g_1)[\omega(h)\omega_p^{-1}(g_1,h)\omega_p^{-1}(g_2^{-1},g_1h)\omega_p(g_2^{-1}g_1,h)\omega^{-1}(h)]\langle 0|T(g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2)$$
$$=\omega(g_1)\omega_p^{-1}(g_2^{-1},g_1)\langle 0|T(g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2)$$
$$=\omega(g_1)\bar{\omega}_p(g_2^{-1},g_1)\langle 0|T(g_1^{-1}g_2)|0\rangle\bar{\omega}(g_2),$$

where, in a very similar way to before, the second to sixth factors in the third line have been contracted via the projective representation associativity condition to give $\omega_p^{-1}(g_2^{-1},g_1)$ in the fourth line. Thus, as implied by our notation, the inner product does not depend specifically on g_1 but just on the coset $x_1=x(g_1)$, and this can similarly be shown for g_2. If all coherent states are operated on by the same $T(g)$, then the inner product of two of the new states, using Eq. (6), is

$$\langle x(gg_1)|x(gg_2)\rangle=\theta(g,x_1)\langle x_1|T^{-1}(g)T(g)|x_2\rangle\bar{\theta}(g,x_2)=\theta(g,x_1)\langle x_1|x_2\rangle\bar{\theta}(g,x_2).$$

Let us now assume that the invariant measure dg on the group induces an invariant measure dx on the set of cosets $M=G/H$. Given sufficient convergence, consider the operator

$$B=\int|x\rangle\langle x|dx;$$

from the definition of B and the invariance of the measure, Eq. (6) implies that

$$T(g)BT^{-1}(g)=\int T(g)|y\rangle\langle y|T(g^{-1})\,dy=\int |x(gy)\rangle\theta(g,y)\bar{\theta}(g,y)\langle x(gy)|\,dy$$

$$=\int |x\rangle\langle x|\,dx=B,$$

so B commutes with all of the operators $T(g)$. By the quaternionic generalization of Schur's Lemma,[5] this means that B is of the form $B_0 1+B_1 I$, where B_0 and B_1 are real, 1 is the usual identity operator, and I is a unit anti-self-adjoint operator, $I^\dagger=-I$; however, since B is clearly self-adjoint, B_1 must vanish, so B is a multiple of the identity as in the compex case. Given a coherent state $|y\rangle$ that is normalized, $\langle y|y\rangle=1$,

$$B_0=\langle y|B|y\rangle=\int \langle y|x\rangle\langle x|y\rangle\,dx=\int |\langle y|x\rangle|^2\,dx=\int |\langle 0|x\rangle|^2\,dx;$$

hence,

$$\frac{1}{B_0}\int |x\rangle\langle x|\,dx=1.$$

With this form of the identity, an arbitrary state may be expanded over the coherent states,

$$|\psi\rangle=\frac{1}{B_0}\int |x\rangle\langle x|\psi\rangle\,dx=\frac{1}{B_0}\int |x\rangle\psi(x)\,dx, \tag{7}$$

where

$$\psi(x)\equiv\langle x|\psi\rangle.$$

Then,

$$\langle\psi|\psi\rangle=\frac{1}{B_0^2}\int\int \bar{\psi}(x)\langle x|y\rangle\psi(y)\,dx\,dy;$$

however,

$$\psi(x)=\langle x|\psi\rangle=\langle x|\,\frac{1}{B_0}\int |y\rangle\langle y|\psi\rangle\,dy=\frac{1}{B_0}\int \langle x|y\rangle\psi(y)\,dy, \tag{8}$$

so,

$$\langle\psi|\psi\rangle=\frac{1}{B_0}\int |\psi(x)|^2\,dx.$$

Defining

$$K(x,y)=\frac{1}{B_0}\langle x|y\rangle,$$

Eq. (8) implies that this is a reproducing kernel,

$$K(x,z)=\int K(x,y)K(y,z)\,dy,$$

and the function

$$\hat{f}(x)=\int K(x,y)f(y)\,dy$$

satisfies Eq. (8), in the place of $\psi(x)$, for an arbitrary function $f(x)$. If $|\psi\rangle$ is itself a coherent state $|y\rangle$, then, from Eq. 7,

$$|y\rangle=\frac{1}{B_0}\int |x\rangle\langle x|y\rangle\,dx,$$

so the coherent states are not linearly independent, meaning that the system of coherent states is overcomplete.

III. THE CASE OF THE SPECIAL NILPOTENT OR WEYL GROUP

Having extended Perolomov's formulation of coherent states to a quaternionic Hilbert space, we continue to follow his paper and consider the case of the nilpotent group. In the complex case, this group leads to the familiar coherent states widely used in quantum optics. The special nilpotent or Weyl group is generated by a set of annihilation operators, $\{a_i\}$, where i runs from 1 to N, their conjugate creation operators, $\{a_i^\dagger\}$, and the identity operator, 1. The commutation relations between these operators are

$$[a_i,a_j]=[a_i^\dagger,a_j^\dagger]=[a_i,1]=[a_i^\dagger,1]=0$$

and

$$[a_i,a_j^\dagger]=\delta_{ij}1.$$

Let the E_A, where A runs from 1 to 3, be three quaternion imaginary operators[6] with an algebra isomorphic to the algebra of i, j, and k and all of which commute with the a_i and the $a_i^\dagger$. Then, an anti-self-adjoint[7] element of the Lie algebra of the group may be written

$$t+\sum_i \beta_i a_i-\sum_i \bar{\beta}_i a_i^\dagger,$$

where t is a quaternion imaginary operator,

$$t=\sum_A t_A E_A,$$

and the β_i are quaternion operators,

$$\beta_i=\beta_{i0}1+\sum_A \beta_{iA}E_A.$$

For convenience, the generator is written using a shorter notation,

$$t+\beta a-\bar{\beta}a^\dagger.$$

The group is then obtained from the algebra by means of the exponential mapping, so that for a general element $g \in G, T(g)$ may be written

$$T(g)=(t,\beta)=\exp(t+\beta a-\bar{\beta}a^{\dagger}).$$

In the case of a complex Hilbert space, as considered by Perelomov, group closure follows quickly. However, in the quaternionic case, t and the β_i may be noncommutative, and requiring group closure will impose significant restrictions. Consider, then, the product of two group elements,

$$(s,\alpha)(t,\beta)=\exp(s+\alpha a-\bar{\alpha}a^{\dagger})\exp(t+\beta a-\bar{\beta}a^{\dagger});$$

using the Baker–Campbell–Hausdorff formula to second order,

$$\exp X \exp Y=\exp(X+Y+\tfrac{1}{2}[X,Y]+\cdots),$$

the product may be written

$$\exp\{s+t+(\alpha+\beta)a-(\bar{\alpha}+\bar{\beta})a^{\dagger}+\tfrac{1}{2}[s+\alpha a-\bar{\alpha}a^{\dagger},t+\beta a-\bar{\beta}a^{\dagger}]\},$$

and so, to obtain a group, this requires that

$$\tfrac{1}{2}[s+\alpha a-\bar{\alpha}a^{\dagger},t+\beta a-\bar{\beta}a^{\dagger}]=u+\gamma a-\bar{\gamma}a^{\dagger}. \tag{9}$$

In particular, the coefficients of $a_i a_j$ and $a_i^{\dagger}a_j^{\dagger}$ must vanish, which requires that

$$[\alpha_i,\beta_j]=0$$

for each i and j; hence, all α_i and β_i must belong to the same $\mathscr{C}(1, I)$ subalgebra rather than being free to range over any quaternion. With this constraint, the coefficients of $a_i a_j^{\dagger}$ and $a_i^{\dagger}a_j$ also vanish, and Eq. (9) implies that

$$u=\tfrac{1}{2}[s,t], \quad \gamma_i=\tfrac{1}{2}[s,\beta_i]+\tfrac{1}{2}[\alpha_i,t].$$

However, the γ_i must have the same structure as the α_i and β_i and thus belong to the same $\mathscr{C}(1, I)$ subalgebra, which requires that s and t are simply proportional to I—consequently, u and the γ_i vanish. Therefore, for group closure, the representation can only be $\mathscr{C}(1, I)$ embedded in the quaternionic Hilbert space rather than fully quaternionic. For the case of the nilpotent group, then, there is no quaternionic generalization of standard complex coherent states.

IV. THE CASE OF INTRINSICALLY QUATERNIONIC IRREDUCIBLE REPRESENTATIONS OF SU(2)

We consider now the anti-self-adjoint generators of SU(2), S_x, S_y and S_z, such that[8]

$$[S_l,S_m]=\sum_n \epsilon_{lmn}S_n.$$

It can readily be observed that a quaternionic realization of this algebra is

$$S_x=\tfrac{1}{2}i, \quad S_y=\tfrac{1}{2}j, \quad S_z=\tfrac{1}{2}k,$$

which is a one-dimensional quaternionic irreducible representation of SU(2). Consider eigenstates of S_z; these can be chosen to be

$$|\tfrac{1}{2}\rangle=1, \quad |-\tfrac{1}{2}\rangle=j$$

such that

$$S_z|\pm\tfrac{1}{2}\rangle=\pm\tfrac{1}{2}|\pm\tfrac{1}{2}\rangle k.$$

Then choosing either of these states as $|0\rangle$, the stationary subgroup is

$$H=\{\exp \alpha S_z\}.$$

Following Perelomov,[1] a coherent state based on such a $|0\rangle$ may be characterized by a vector $\mathbf{n}$ or by a polar angle θ and an azimuthal angle ϕ. For the purposes of an example, choose $|0\rangle$ to be $|\tfrac{1}{2}\rangle$ and then

$$|\mathbf{n}\rangle=\exp \phi S_z \exp \theta S_y|0\rangle=\exp \tfrac{1}{2}\phi k \exp \tfrac{1}{2}\theta j;$$

then

$$\langle\mathbf{n}'|\mathbf{n}\rangle=\exp-\tfrac{1}{2}\theta' j \exp \tfrac{1}{2}(\phi-\phi')k \exp \tfrac{1}{2}\theta j$$

and hence

$$|\langle 0|\mathbf{n}\rangle|^2=1$$

so that

$$B_0=\int |\langle 0|\mathbf{n}\rangle|^2 \mathbf{dn}=\mathbf{4}\pi.$$

Correspondingly,

$$\frac{1}{4\pi}\int |\mathbf{n}\rangle\langle\mathbf{n}|\, d\mathbf{n}=1,$$

as expected.

Since, with the above definitions for the S_n,

$$S^2=-S_x^2-S_y^2-S_z^2=\tfrac{3}{4},$$

this can be seen to be a one-dimensional spin $\tfrac{1}{2}$ representation of SU(2). This is a special case of a general result due to Finkelstein *et al.*[9] which states that besides the real, integer spin (Frobenius–Schur class $+1$) and the half-integer spin (Frobenius–Schur class -1) representations, there are quaternionic representations for half-integer spin of precisely half the dimension of the Frobenius–Schur class -1 representations for the same spin. For instance, one choice for the spin $\tfrac{3}{2}$ representation is

$$S_x=\frac{1}{2}i\begin{pmatrix} 2 & \sqrt{3}j \\ \sqrt{3}j & 0 \end{pmatrix}, \quad S_y=\frac{1}{2}j\begin{pmatrix} 2 & -\sqrt{3}j \\ \sqrt{3}j & 0 \end{pmatrix}, \quad S_z=\frac{1}{2}k\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Then,

$$S^2=-S_x^2-S_y^2-S_z^2=\frac{15}{4}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so the spin is indeed $\frac{3}{2}$. Consider, as before, eigenstates of S_z; these may be chosen to be

$$\left|\frac{3}{2}\right\rangle=\begin{pmatrix}0\\1\end{pmatrix},\quad \left|\frac{1}{2}\right\rangle=\begin{pmatrix}1\\0\end{pmatrix},\quad \left|-\frac{1}{2}\right\rangle=\begin{pmatrix}j\\0\end{pmatrix},\quad \left|-\frac{3}{2}\right\rangle=\begin{pmatrix}0\\j\end{pmatrix}.$$

Again, when one of these is chosen as $|0\rangle$, the stationary subgroup is

$$H=\{\exp \alpha S_z\}.$$

As an example, choose $|0\rangle$ to be $|\frac{3}{2}\rangle$ so that

$$\begin{aligned}
|\mathbf{n}\rangle&=\exp \phi S_z \exp \theta S_y|0\rangle\\
&=\exp\frac{1}{2}\,\phi\begin{pmatrix}k&0\\0&3k\end{pmatrix}\exp\frac{1}{2}\,\theta\begin{pmatrix}2j&\sqrt{3}\\-\sqrt{3}&0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\\
&=\exp\frac{1}{2}\,\phi\begin{pmatrix}k&0\\0&3k\end{pmatrix}\exp\frac{1}{2}\,\theta\begin{pmatrix}2j&\sqrt{3}\\-\sqrt{3}&0\end{pmatrix}\frac{3}{4}\left[-\begin{pmatrix}1\\ \frac{j}{\sqrt{3}}\end{pmatrix}\frac{j}{\sqrt{3}}+\begin{pmatrix}\frac{j}{\sqrt{3}}\\1\end{pmatrix}\right]\\
&=\exp\frac{1}{2}\,\phi\begin{pmatrix}k&0\\0&3k\end{pmatrix}\frac{3}{4}\left[-\begin{pmatrix}1\\ \frac{j}{\sqrt{3}}\end{pmatrix}\frac{j}{\sqrt{3}}\exp\frac{3}{2}\,\theta j+\begin{pmatrix}\frac{j}{\sqrt{3}}\\1\end{pmatrix}\exp-\frac{1}{2}\,\theta j\right]\\
&=\frac{3}{4}\left[-\begin{pmatrix}\exp\frac{1}{2}\phi k\\ \exp\frac{3}{2}\,\phi k\,\frac{j}{\sqrt{3}}\end{pmatrix}\frac{j}{\sqrt{3}}\exp\frac{3}{2}\,\theta j+\begin{pmatrix}\exp\frac{1}{2}\,\phi k\,\frac{j}{\sqrt{3}}\\ \exp\frac{3}{2}\,\phi k\end{pmatrix}\exp-\frac{1}{2}\,\theta j\right]\\
&=\frac{3}{4}\begin{pmatrix}\exp\frac{1}{2}\,\phi k\left[-\exp\frac{3}{2}\,\theta j+\exp-\frac{1}{2}\,\theta j\right]\frac{j}{\sqrt{3}}\\ \exp\frac{3}{2}\,\phi k\left[\frac{1}{3}\exp\frac{3}{2}\,\theta j+\exp-\frac{1}{2}\,\theta j\right]\end{pmatrix},
\end{aligned}$$

where the calculation has been carried out efficiently by decomposing $\binom{0}{1}$ into S_y eigenstates. From this, we find

$$\begin{aligned}
\langle\mathbf{n}'|\mathbf{n}\rangle=\tfrac{9}{16}\Big[&\tfrac{1}{3}\Big(-\exp-\tfrac{3}{2}\,\theta' j+\exp\tfrac{1}{2}\,\theta' j\Big)\exp\tfrac{1}{2}\,(\phi'-\phi)k\Big(-\exp\tfrac{3}{2}\,\theta j+\exp-\tfrac{1}{2}\,\theta j\Big)\\
&+\Big(\tfrac{1}{3}\exp-\tfrac{3}{2}\,\theta' j+\exp\tfrac{1}{2}\,\theta' j\Big)\exp\tfrac{3}{2}\,(\phi-\phi')k\Big(\tfrac{1}{3}\exp\tfrac{3}{2}\,\theta j+\exp-\tfrac{1}{2}\,\theta j\Big)\Big],
\end{aligned}$$

and hence

$$|\langle 0|\mathbf{n}\rangle|^2=\tfrac{5}{8}+\tfrac{3}{8}\cos 2\theta,$$

so that

$$B_0=\int |\langle 0|\mathbf{n}\rangle|^2\, d\mathbf{n}=2\pi.$$

Correspondingly,

$$\frac{1}{2\pi}\int |\mathbf{n}\rangle\langle\mathbf{n}|\, d\Omega=\frac{3}{8\pi}\int\begin{pmatrix}\sin^2\theta & x(\theta,\phi)\\ \bar{x}(\theta,\phi) & \frac{1}{3}+\cos^2\theta\end{pmatrix}d\Omega$$

with

$$x(\theta,\phi)=\frac{1}{\sqrt{3}}\,(j\sin^2\theta e^{-2\phi k}+\sin 2\theta e^{-\phi k}),$$

which on doing the ϕ integration becomes

$$\frac{3}{4}\int_0^\pi\begin{pmatrix}\sin^2\theta & 0\\ 0 & \frac{1}{3}+\cos^2\theta\end{pmatrix}\sin\theta\, d\theta=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},$$

as expected. These examples can be readily extended to the general half-integral spin quaternionic irreducible representations of SU(2).

ACKNOWLEDGMENTS

SLA wishes to thank J. R. Klauder for stimulating discussions about quaternionic coherent states and to acknowledge the DOE for its support under Grant No. DE-FG02-90ER40542.

[1] A. M. Perelomov, "Coherent States for Arbitrary Lie Group," Commun. Math. Phys. **26**, 222–235 (1972). See also J. R. Klauder, J. Math. Phys. **4**, 1058–1073 (1963), Sec. 3, for earlier work on generalized coherent states.
[2] S. L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford U.P., New York, 1995).
[3] See S. L. Adler, "Projective Group Representations in Quaternionic Hilbert Space," J. Math. Phys. **37**, 2352–2360 (1996) and T. Tao and A. C. Millard, "On the Structure of Projective Group Representations in Quaternionic Hilbert Space," J. Math. Phys. **37**, 5848–5857 (1996).
[4] J. R. Klauder, private communication to S. L. Adler.
[5] This is due to Emch; an exposition of it, and references to Emch's papers on it, may be found in Ref. 2.
[6] These are also represented by I, J, and K, especially when a specific form in terms of i, j, and k is given. Reference 2 deals with their construction in more detail.
[7] See Ref. 2 for a discussion of why anti-self-adjoint, rather than self-adjoint, operators are of interest.
[8] The sign convention for these operators is the opposite to that used in Ref. 2.
[9] D. Finkelstein, J. M. Jauch, and D. Speiser, "Notes on Quaternion Quantum Mechanics" (1959), in *Logico-Algebraic Approach to Quantum Mechanics II*, edited by C. Hooker (Reidel, Dordrecht, 1979).

Projective group representations in quaternionic Hilbert space

Stephen L. Adler[a)]
Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540

(Received 16 January 1996; accepted for publication 26 January 1996)

We extend the discussion of projective group representations in quaternionic Hilbert space that was given in our recent book. The associativity condition for quaternionic projective representations is formulated in terms of unitary operators and then analyzed in terms of their generator structure. The multi-centrality and centrality assumptions are also analyzed in generator terms, and implications of this analysis are discussed. © *1996 American Institute of Physics.*
[S0022-2488(96)01105-7]

I. ASSOCIATIVITY CONDITION FOR QUATERNIONIC PROJECTIVE GROUP REPRESENTATIONS

In quaternionic quantum mechanics, all symmetries of the transition probabilities are generated by unitary transformations acting on the states of Hilbert space.[1-3] In the simplest case, the unitary transformations $U_a, U_b,\ldots$ form a representation (or vector representation) of the symmetry group with elements $a,b,\ldots,$

$$U_b U_a = U_{ba}. \tag{1}$$

A more general possibility is that the group multiplication table is represented over the *rays* corresponding to a complete set of physical states, but not over individual state vectors chosen as ray representatives. This more general composition rule defines a quaternionic *projective representation* (or ray representation), and takes the form (Ref. 4, Sec. 4.3)

$$U_b U_a|f\rangle = U_{ba}|f\rangle\omega(f;b,a), \quad |\omega(f;b,a)|=1, \tag{2}$$

for one particular complete set of states $|f\rangle$ and a set of quaternionic phases $\omega(f;b,a)$. When we change ray representative from $|f\rangle$ to $|f_\phi\rangle \equiv |f\rangle\phi$, with $|\phi|=1$, the phase defining the projective representation is easily seen to transform as

$$\omega(f_\phi;b,a) = \bar{\phi}\omega(f;b,a)\phi, \tag{3}$$

with the bar denoting quaternion conjugation. Equation (3) shows clearly that the projective phase ω must depend on the state label f as well as on the group elements a,b; failure to take this into account can lead[4] to erroneous conclusions (as in Ref. 5) concerning quaternionic projective representations.

The defining relation for quaternionic projective representations given in Eq. (2) can be rewritten in operator form by defining a left-acting operator $\Omega(b,a)$,

$$\Omega(b,a) = \sum_f |f\rangle\omega(f;b,a)\langle f|, \tag{4a}$$

[a)]Electronic mail address: adler@sns.ias.edu

which, using Eq. (3), is seen to be independent of the ray representative chosen for the states $|f\rangle$. Multiplying Eq. (2) from the right by $\langle f|$ and summing over the complete set of states $|f\rangle$, we obtain the operator form of the projective representation,

$$U_b U_a = U_{ba} \Omega(b,a). \tag{4b}$$

It is also immediate from the definition of Eq. (4a), and the fact that $|\omega|=1$, that the operator $\Omega(b,a)$ is quaternion unitary,

$$\Omega(b,a)^\dagger \Omega(b,a) = \Omega(b,a) \Omega(b,a)^\dagger = 1. \tag{5}$$

Note that if we were to make the definition of a quaternionic projective representation more restrictive by requiring that Eq. (2) hold for *all* states in Hilbert space, rather than for one particular complete set of states, then we would require $\Omega(b,a)=1$, since the unit operator is the only unitary operator which is simultaneously diagonal on all complete bases in quaternionic Hilbert space. Hence this more restrictive definition excludes quaternionic embeddings of complex projective representations, whereas these are admitted as quaternionic projective representations by the definition of Eq. (2).

A nontrivial condition on the projective representation structure is obtained from the associativity of multiplication in quaternionic Hilbert space, which implies

$$(U_c U_b) U_a = U_c (U_b U_a). \tag{6}$$

Applying Eq. (4b) twice to the left-hand side of Eq. (6), we obtain

$$(U_c U_b) U_a = U_{cb} \Omega(c,b) U_a = U_{cb} U_a U_a^{-1} \Omega(c,b) U_a = U_{cba} \Omega(cb,a) U_a^{-1} \Omega(c,b) U_a, \tag{7a}$$

while applying Eq. (4b) twice to the right-hand side of Eq. (6) gives

$$U_c (U_b U_a) = U_c U_{ba} \Omega(b,a) = U_{cba} \Omega(c,ba) \Omega(b,a). \tag{7b}$$

Upon multiplying from the left by U_{cba}^{-1}, Eqs. (7a) and (7b) give the operator form of the *associativity condition*:

$$\Omega(c,ba) \Omega(b,a) = \Omega(cb,a) U_a^{-1} \Omega(c,b) U_a. \tag{8}$$

We can also express the associativity condition as a condition on the quaternionic phase $\omega(f;b,a)$ introduced in Eq. (2), by applying the spectral representation of Eq. (4a) to the operator form of the associativity condition given in Eq. (8). From Eq. (4a) we obtain

$$\Omega(c,ba) = \sum_f |f\rangle \omega(f;c,ba) \langle f|, \tag{9a}$$

which when multiplied from the right by Eq. (4a) gives

$$\Omega(c,ba) \Omega(b,a) = \sum_f |f\rangle \omega(f;c,ba) \omega(f;b,a) \langle f|. \tag{9b}$$

Equation (4a) and the unitarity of $\Omega(cb,a)$ also imply that

$$\Omega(cb,a)^{-1} = \sum_f |f\rangle \overline{\omega(f;cb,a)} \langle f|, \tag{9c}$$

and so the associativity condition of Eq. (8) can be rewritten as

$$U_a^{-1}\Omega(c,b)U_a=\Omega(cb,a)^{-1}\Omega(c,ba)\Omega(b,a)=\sum_f |f\rangle\overline{\omega(f;cb,a)}\omega(f;c,ba)\omega(f;b,a)\langle f|. \tag{10}$$

Hence $U_a^{-1}\Omega(c,b)U_a$ is diagonal in the basis spanned by the states $|f\rangle$. Taking matrix elements of Eq. (10), and using the unitarity of U_a, the associativity condition gives the two relations

$$\overline{\omega(f;cb,a)}\omega(f;c,ba)\omega(f;b,a)=\sum_{f''}\overline{\langle f''|U_a|f\rangle}\omega(f'';c,b)\langle f''|U_a|f\rangle, \tag{11}$$

and, when $\langle f'|f\rangle=0$,

$$0=\sum_{f''}\overline{\langle f''|U_a|f\rangle}\omega(f'';c,b)\langle f''|U_a|f'\rangle. \tag{12}$$

We conclude this section by comparing the quaternionic Hilbert space form of the associativity condition with the simpler form which is familiar from complex Hilbert space.[6,7] In a complex Hilbert space, the phase $\omega(f;b,a)$ introduced in Eq. (2) is a complex number, and commutes with the phase ϕ, also now complex, which we introduced in Eq. (3) to describe a change of ray representative. Hence Eq. (3) implies, in the complex case, that $\omega(f;b,a)$ is independent of the ray representative chosen for the state $|f\rangle$, and it is then consistent to assume that $\omega(f;b,a)$ is independent of the state label f, so that

$$\omega(f;b,a)=\omega(b,a) \quad \text{complex Hilbert space.} \tag{13a}$$

Substituting Eq. (13a) into Eq. (4a), we now obtain

$$\Omega(b,a)=\sum_f |f\rangle\omega(b,a)\langle f|=\omega(b,a)\sum_f |f\rangle\langle f|=\omega(b,a)1, \tag{13b}$$

where 1 denotes the unit operator in complex Hilbert space. Since the complex phase $\omega(b,a)$ is a c-number in complex Hilbert space, on substituting Eq. (13b) into Eq. (4b) we learn that

$$U_bU_a=U_{ba}\omega(b,a)=\omega(b,a)U_{ba}, \tag{14a}$$

which is the standard definition of a projective representation in complex Hilbert space. Moreover, since Eq. (13b) implies that $\Omega(b,a)$ commutes with the unitary operator U_a, the associativity condition of Eqs. (8) and (11) reduces to the familiar complex Hilbert space form

$$\omega(c,ba)\omega(b,a)=\omega(cb,a)\omega(c,b). \tag{14b}$$

II. THE ASSOCIATIVITY CONDITION IN GENERATOR FORM

Let us now assume that the symmetry group with which we are dealing is a Lie group, so that in the neighborhood of the identity e the unitary transformations $U_a, U_b, U_{ba}, \ldots$ can be written in terms of a set of anti-self-adjoint generators $\tilde{G}_A$ as

$$U_a=\exp\left(\sum_A \theta_A^a\tilde{G}_A\right), \quad U_b=\exp\left(\sum_A \theta_A^b\tilde{G}_A\right), \quad U_{ba}=\exp\left(\sum_A \theta_A^{ba}\tilde{G}_A\right),\ldots, \tag{15a}$$

with $\theta_A^e=0$ and $U_e=1$. Then Eq. (4b) implies that $\Omega(b,a)$ must be unity when either a or b is the identity, and thus the generator form for this operator is

$$\Omega(b,a)=\exp\left(\frac{1}{2}\sum_{BA}\left[\theta_B^b\theta_A^a\tilde{I}_{BA}+\sum_C\theta_B^b\theta_C^b\theta_A^a\tilde{J}^{(1)}_{(BC)A}+\sum_C\theta_B^b\theta_A^a\theta_C^a\tilde{J}^{(2)}_{B(AC)}+O(\theta^4)\right]\right), \quad (15b)$$

where the parentheses ( ) around a set of indices indicate that the tensor in question is symmetric in those indices, and where we use the tilde to indicate operators which are anti-self-adjoint. The parameters θ_C^{ba} must be functions of the parameters θ_A^a and θ_B^b,

$$\theta_C^{ba}=\psi_C^{ba}(\{\theta_B^b\},\{\theta_A^a\})=\theta_C^b+\theta_C^a+\frac{1}{2}\sum_{BA}C_{BAC}\theta_B^b\theta_A^a+O(\theta^3), \quad (15c)$$

where in making the Taylor expansion we have used the fact that $U_{be}=U_b$ and $U_{ea}=U_a$, which fixes the linear terms in the expansion and requires the quadratic term to be bilinear.

We proceed now to derive a number of relations by combining the generator expansions of Eqs. (15a)–(15c) with the formulas of Sec. I. We begin by substituting Eqs. (15a)–(15c) into Eq. (4b) using the Baker–Campbell–Hausdorff formula,

$$\exp X \exp Y=\exp(X+Y+\tfrac{1}{2}[X,\ Y]+\cdots), \quad (16a)$$

to combine exponents arising from the factors on the left and right. From the left-hand side of Eq. (4b) we obtain,

$$U_bU_a=\exp\left(\sum_B\theta_B^b\tilde{G}_B+\sum_A\theta_A^a\tilde{G}_A+\frac{1}{2}\sum_{BA}\theta_B^b\theta_A^a[\tilde{G}_B,\ \tilde{G}_A]+O(\theta^3)\right), \quad (16b)$$

while from the right-hand side of Eq. (4b) we obtain

$$U_{ba}\Omega(b,a)=\exp\left(\sum_C(\theta_C^b+\theta_C^a)\tilde{G}_C+\frac{1}{2}\sum_{CBA}C_{BAC}\theta_B^b\theta_A^a\tilde{G}_C+\frac{1}{2}\sum_{BA}\theta_B^b\theta_A^a\tilde{I}_{BA}+O(\theta^3)\right). \quad (16c)$$

Equating Eqs. (16b) and (16c) thus gives the relations

$$[\tilde{G}_B,\ \tilde{G}_A]=\sum_C C_{[BA]C}\tilde{G}_C+\tilde{I}_{[BA]} \quad (17a)$$

and

$$0=\sum_C C_{(BA)C}\tilde{G}_C+\tilde{I}_{(BA)}, \quad (17b)$$

where the square brackets [ ] around a set of indices indicates that the tensor in question is antisymmetric in these indices. We shall restrict ourselves henceforth to the case in which $C_{(BA)C}=0$, which by Eq. (17b) implies that $\tilde{I}_{(BA)}=0$; making this assumption then implies that $C_{BAC}=C_{[BA]C}$ and $\tilde{I}_{BA}=\tilde{I}_{[BA]}$. In other words, we are assuming that the structure constants C_{BAC} for a projective representation have the same antisymmetric form as holds for a vector representation. Changing the summation index C to D in Eq. (17a), and then taking the commutator of Eq. (17a) with $\tilde{G}_C$, we find

$$[\tilde{G}_C,\ [\tilde{G}_B,\ \tilde{G}_A]]=\sum_D C_{[BA]D}[\tilde{G}_C,\ \tilde{G}_D]+[\tilde{G}_C,\ \tilde{I}_{[BA]}]; \quad (18a)$$

adding to this identity the two related identities obtained by cyclically permuting A,B,C, using the fact that the left-hand side of the sum vanishes by the Jacobi identity for the commutator, and substituting Eq. (17a) for the commutators appearing on the right-hand side of the sum, we obtain the identity

$$\begin{aligned}&\sum_{DE} (C_{[BA]D}C_{[CD]E}+C_{[CB]D}C_{[AD]E}+C_{[AC]D}C_{[BD]E})\tilde{G}_E\\&+\sum_{D} (C_{[BA]D}\tilde{I}_{[CD]}+C_{[CB]D}\tilde{I}_{[AD]}+C_{[AC]D}\tilde{I}_{[BD]})\\&+[\tilde{G}_C,\tilde{I}_{[BA]}]+[\tilde{G}_A,\tilde{I}_{[CB]}]+[\tilde{G}_B,\tilde{I}_{[AC]}]=0.\end{aligned}\tag{18b}$$

We next substitute Eqs. (15a)–(15c) into the associativity condition of Eq. (8), now keeping cubic terms in the exponent of the form $\theta_A^a\theta_B^b\theta_C^c$, but dropping cubic terms, such as $\theta_A^a\theta_B^a\theta_C^c$, that do not contain all three of the upper indices a,b,c. For the first factor on the left-hand side of Eq. (8), we find from Eqs. (15b) and (15c) that

$$\begin{aligned}\Omega(c,ba)&=\exp\left(\frac{1}{2}\sum_{BA}\left(\theta_B^c\theta_A^{ba}\tilde{I}_{[BA]}+\sum_C \theta_B^c\theta_A^{ba}\theta_C^{ba}\tilde{J}^{(2)}_{B(AC)}\right)\right)\\&=\exp\left(\frac{1}{2}\sum_{BA}\left[\theta_B^c\left(\theta_A^b+\theta_A^a+\frac{1}{2}\sum_{DE} C_{[DE]A}\theta_D^b\theta_E^a\right)\tilde{I}_{[BA]}\right.\right.\\&\left.\left.+2\sum_C \theta_B^c\theta_A^b\theta_C^a\tilde{J}^{(2)}_{B(AC)}\right]\right),\end{aligned}\tag{19a}$$

while for the second factor on the left-hand side of Eq. (8) we have

$$\Omega(b,a)=\exp\left(\frac{1}{2}\sum_{BA}\theta_B^b\theta_A^a\tilde{I}_{[BA]}\right).\tag{19b}$$

Since the exponents in Eqs. (19a) and (19b) both begin at order θ^2, through order θ^3 we can simply add exponents to get the product on the left-hand side of Eq. (8). Proceeding similarly for the first factor on the right-hand side of Eq. (8), we obtain

$$\begin{aligned}\Omega(cb,a)&=\exp\left(\frac{1}{2}\sum_{BA}\left(\theta_B^{cb}\theta_A^a\tilde{I}_{[BA]}+\sum_C \theta_B^{cb}\theta_C^{cb}\theta_A^a\tilde{J}^{(1)}_{(BC)A}\right)\right)\\&=\exp\left(\frac{1}{2}\sum_{BA}\left[\left(\theta_B^c+\theta_B^b+\frac{1}{2}\sum_{DE} C_{[DE]B}\theta_D^c\theta_E^b\right)\theta_A^a\tilde{I}_{[BA]}\right.\right.\\&\left.\left.+2\sum_C \theta_B^c\theta_C^b\theta_A^a\tilde{J}^{(1)}_{(BC)A}\right]\right),\end{aligned}\tag{20a}$$

while for the second factor on the right-hand side of Eq. (8), use of the Baker–Campbell–Hausdorff formula gives

$$U_a^{-1}\Omega(c,b)U_a=\exp\left(-\sum_A \theta_A^a\tilde{G}_A\right)\exp\left(\frac{1}{2}\sum_{CB}\theta_C^c\theta_B^b\tilde{I}_{[CB]}\right)\exp\left(\sum_A \theta_A^a\tilde{G}_A\right)$$

$$=\exp\left(\frac{1}{2}\sum_{CB}\theta_C^c\theta_B^b\tilde{I}_{[CB]}-\frac{1}{2}\sum_{A}\sum_{CB}\theta_A^a\theta_C^c\theta_B^b[\tilde{G}_A,\tilde{I}_{[CB]}]\right). \tag{20b}$$

Since the exponents in Eqs. (20a) and (20b) begin at order θ^2, it again suffices to simply add the exponents to form the product appearing on the right-hand side of Eq. (8). Thus, to the requisite order, the content of Eq. (8) is obtained by equating the sum of the exponents in Eqs. (19a) and (19b) to the corresponding sum of exponents in Eqs. (20a) and (20b). The quadratic terms in θ are immediately seen to be identical on left and right, while the cubic term proportional to $\theta_A^a\theta_B^b\theta_C^c$ gives (after some relabeling of dummy summation indices) the nontrivial identity

$$\tilde{J}^{(2)}_{C(BA)}+\frac{1}{4}\sum_D C_{[BA]D}\tilde{I}_{[CD]}=\tilde{J}^{(1)}_{(CB)A}+\frac{1}{4}\sum_D C_{[CB]D}\tilde{I}_{[DA]}-\frac{1}{2}[\tilde{G}_A,\tilde{I}_{[CB]}]. \tag{21}$$

On totally antisymmetrizing with respect to the indices A,B,C, the terms in Eq. (21) involving $\tilde{J}^{(1,2)}$ drop out, and we are left with the identity

$$\sum_D (C_{[BA]D}\tilde{I}_{[CD]}+C_{[CB]D}\tilde{I}_{[AD]}+C_{[AC]D}\tilde{I}_{[BD]})+[\tilde{G}_C,\tilde{I}_{[BA]}]+[\tilde{G}_A,\tilde{I}_{[CB]}]+[\tilde{G}_B,\tilde{I}_{[AC]}]=0. \tag{22a}$$

In other words, associativity implies that the sum of the second and third lines of Eq. (18b) vanishes separately; hence the first line of Eq. (18b) must also vanish, and since the generators $\tilde{G}_E$ are linearly independent this gives the Jacobi identity for the structure constants,

$$\sum_{DE}(C_{[BA]D}C_{[CD]E}+C_{[CB]D}C_{[AD]E}+C_{[AC]D}C_{[BD]E})=0. \tag{22b}$$

In the complex case, in which $\Omega(a,b)=\omega(a,b)1$ is a c-number, the tensor $\tilde{I}_{[AB]}$ is a c-number "central charge" and the commutator terms in Eqs. (18b) and (22a) vanish identically. Therefore, in the complex case, Eq. (18b) implies both Eq. (22b) and the identity

$$\sum_D (C_{[BA]D}\tilde{I}_{[CD]}+C_{[CB]D}\tilde{I}_{[AD]}+C_{[AC]D}\tilde{I}_{[BD]})=0 \quad \text{complex case,} \tag{23}$$

and so one obtains the entire content of the associativity condition from the simpler analysis leading to Eq. (18b), without having to perform the third-order expansion needed to obtain Eq. (22a).

III. GENERAL, MULTI-CENTRAL, AND CENTRAL QUATERNIONIC PROJECTIVE REPRESENTATIONS

The analysis of Sec. II applies to the general case (apart from the restriction $C_{(BA)C}=0$) of a quaternionic projective representation; in order to obtain more detailed results it is necessary to introduce further structural assumptions. In Ref. 4 two special classes of quaternionic projective representations are defined. A quaternionic projective representation is defined to be *multi-central* if

$$[\Omega(b,a),\ U_a]=[\Omega(b,a),\ U_b]=0, \quad \text{all } a,b, \tag{24a}$$

while it is defined to be *central* if

$$[\Omega(b,a),\ U_c]=0, \quad \text{all } a,b,c. \tag{24b}$$

Expressed in terms of the generators introduced in Eqs. (15a)–(15b), the multi-centrality condition takes the form

$$\sum_{ABC} \theta_A^a \theta_B^b \theta_C^a [\tilde{G}_C, \tilde{I}_{[BA]}] = \sum_{ABC} \theta_A^a \theta_B^b \theta_C^b [\tilde{G}_C, \tilde{I}_{[BA]}] = 0, \quad \text{all } a,b, \tag{25a}$$

while the centrality condition becomes

$$\sum_{ABC} \theta_A^a \theta_B^b \theta_C^c [\tilde{G}_C, \tilde{I}_{[BA]}] = 0, \quad \text{all } a,b,c. \tag{25b}$$

Making the definition

$$\Delta_{[AB]C} = [\tilde{G}_C, \tilde{I}_{[BA]}], \tag{25c}$$

we see from Eq. (25a) that multi-centrality requires that $\Delta_{[AB]C}$ be antisymmetric in A,C and in B,C as well as in A,B; thus in the multi-central case Δ is totally antisymmetric, which we will indicate by writing it as $\Delta_{[ABC]}$. From Eq. (25b), we see that centrality requires that $\Delta_{[AB]C}$ must vanish.

Using the generator formulation, we proceed now to discuss successively the general, multi-central, and central cases in the light of the associativity analysis of Sec. II.

(1) *The general case.* An example given in Eqs. (13.54g) and (14.23a) of Ref. 4 shows that one can have a quaternionic projective representation which is neither central nor multi-central. The example is constructed from n independent fermion creation and annihilation operators $b_\ell^\dagger$, b_ℓ, $\ell=1,\ldots,n$, which commute with a left algebra quaternion basis $E_0=1, E_1=I, E_2=J, E_3=K$. Consider the set of three generators $\tilde{G}_A$ defined by

$$\tilde{G}_A = -\tfrac{1}{2} E_A N, \quad A=1,2,3, \tag{26a}$$

with N the number operator

$$N = \sum_{\ell=1}^{n} b_\ell^\dagger b_\ell. \tag{26b}$$

The commutator algebra of these generators has the form of a projective representation of SU(2),

$$\begin{aligned} [\tilde{G}_B, \tilde{G}_A] &= -\sum_{C=1}^{3} \epsilon_{[BAC]} \tilde{G}_C + \tilde{I}_{[BA]}, \\ \tilde{I}_{[BA]} &= \sum_{C=1}^{3} \epsilon_{[BAC]} \frac{1}{2} E_C N(N-1), \end{aligned} \tag{26c}$$

with ϵ the usual three-index antisymmetric tensor. A simple calculation now shows that

$$[\tilde{G}_A, \tilde{I}_{[BC]}] = -N(N-1)(\delta_{AB}\tilde{G}_C - \delta_{AC}\tilde{G}_B), \tag{27a}$$

which is not antisymmetric in either the index pair A,C or the pair A,B, and so the multi-centrality condition is not satisfied. Another simple calculation shows that

$$\sum_D (\epsilon_{[BAD]}\tilde{I}_{[CD]} + \epsilon_{[CBD]}\tilde{I}_{[AD]} + \epsilon_{[ACD]}\tilde{I}_{[BD]}) = 0, \tag{27b}$$

by virtue of the Jacobi identity for the structure constant ϵ, and also

$$[\tilde{G}_C,\tilde{I}_{[BA]}]+[\tilde{G}_A,\tilde{I}_{[CB]}]+[\tilde{G}_B,\tilde{I}_{[AC]}]=0. \tag{27c}$$

Hence the associativity condition of Eq. (22a) is satisfied, with the first and second lines each vanishing separately.

(2) *The multi-central case.* Let us now consider the multi-central case, in which $\Delta_{[AB]C}$ defined in Eq. (25c) is totally antisymmetric in A,B,C, as indicated by the notation $\Delta_{[ABC]}$. The associativity condition of Eq. (22a) then simplifies to

$$\sum_D (C_{[BA]D}\tilde{I}_{[CD]}+C_{[CB]D}\tilde{I}_{[AD]}+C_{[AC]D}\tilde{I}_{[BD]})+3\Delta_{[ABC]}=0. \tag{28a}$$

A further equation involving Δ is obtained from the Jacobi identity

$$[\tilde{G}_D,[\tilde{G}_C,\tilde{I}_{[BA]}]]-[\tilde{G}_C,[\tilde{G}_D,\tilde{I}_{[BA]}]]=[\tilde{I}_{[BA]},[\tilde{G}_C,\tilde{G}_D]], \tag{28b}$$

which on substituting Eqs. (17a) and (25c) becomes

$$[\tilde{G}_D,\Delta_{[AB]C}]-[\tilde{G}_C,\Delta_{[AB]D}]=-\sum_E C_{[CD]E}\Delta_{[AB]E}+[\tilde{I}_{[BA]},\tilde{I}_{[CD]}], \tag{28c}$$

an equation which holds even in the general case in which Δ is not totally antisymmetric. Specializing Eq. (28c) to the multi-central case and contracting it with $\delta_{AC}\delta_{BD}$, the left-hand side vanishes because of the antisymmetry of Δ, while the commutator term on the right-hand side becomes $\Sigma_{AB}[\tilde{I}_{[BA]},\tilde{I}_{[AB]}]=0$, leaving the identity (after relabeling the dummy index E as C)

$$\sum_{ABC} C_{[AB]C}\Delta_{[ABC]}=0. \tag{29}$$

Thus in order for a multi-central projective representation to exist which has $\Delta\neq0$ and so is not also central, there must be a three-index antisymmetric tensor $\Delta_{[ABC]}$ which vanishes when all three indices are contracted with the structure constant $C_{[AB]C}$. This condition is not easy to satisfy and so we pose the question, which we have not been able to answer: Can one construct an example of a multi-central quaternionic projective representation which is not central, or can one prove (in general, or with a restriction, e.g., to simple or semi-simple groups) that a multi-central quaternionic projective representation must always be central? The application of multi-centrality in Ref. 4 sheds no light on this issue; multi-centrality was used there (e.g., in Sec. 12.3) to show that quaternionic Poincaré group projective representations outside the zero energy sector can always be transformed to complex Poincaré group projective representations, which in the sector continuously connected to the identity are known[8] to be transformable to vector representations.

(3) *The central case.* Let us finally consider the central case in which $\Delta=0$, which by Eqs. (25c) and (28c) implies that $\tilde{I}_{[BA]}$ commutes with both $\tilde{G}_C$ and $\tilde{I}_{[CD]}$ for arbitrary values of the indices. Thus $\tilde{I}_{[BA]}$ behaves as a central charge, justifying the name "central" for this case. The various results obtained in Bargmann[6] can be immediately generalized to the quaternionic central case; for example, the analysis of Ref. 6 can be easily extended to show that the central charges associated with a quaternionic central projective representation of a semi-simple Lie group can always be removed by redefinition of the generators; and again, the nontrivial illustration[6] of a complex projective representation, constructed in terms of the phase space translation generators in nonrelativistic quantum mechanics, can be embedded[4] in quaternionic quantum mechanics as a central projective representation.

ACKNOWLEDGMENTS

This work was supported in part by the Department of Energy under Grant No. DE-FG02-90ER40542. I wish to acknowledge the hospitality of the Aspen Center for Physics and of the Department of Applied Mathematics and Theoretical Physics and Clare Hall at Cambridge University, where parts of this work were done, and wish to thank L. Horwitz and E. Witten for helpful conversations.

[1] G. Emch and C. Piron, J. Math. Phys. **4**, 469 (1963).
[2] U. Uhlhorn, Arkiv Phys. **23**, 307 (1963).
[3] V. Bargmann, J. Math. Phys. **5**, 862 (1964).
[4] S. L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford U.P., New York, 1995); see especially Sec. 4.3.
[5] G. Emch, Helv. Phys. Acta **36**, 739, 770 (1963).
[6] V. Bargmann, Ann. Math. **59**, 1 (1954).
[7] S. Weinberg, *The Quantum Theory of Fields, Vol. 1* (Cambridge U.P., Cambridge, 1995).
[8] E. P. Wigner, Ann. Math. **40**, 149 (1939).

A rejoinder on quaternionic projective representations

S. L. Adler
Institute for Advanced Study, Princeton, New Jersey 08540
G. G. Emch
Department of Mathematics, University of Florida, Gainesville, Florida 32611

(Received 18 April 1997; accepted for publication 22 April 1997)

In a series of papers published in this journal, a discussion was started on the significance of a new definition of projective representations in quaternionic Hilbert spaces. In the present paper we give what we believe is a resolution of the semantic differences that had apparently tended to obscure the issues. © *1997 American Institute of Physics.* [S0022-2488(97)01709-X]

I. WIGNER'S THEOREM REVISITED

We must first harmonize the notations in papers[1–4] that were written more than 30 years apart, and for different audiences. Let $\mathscr{H}_{\mathbb{H}}$ be a quaternionic Hilbert space. In order to facilitate the transcription to Dirac's bra-ket notation, we write the multiplication by scalars on the right, with the scalar product defined to be linear in its second term:

$$(\psi p,\phi q)=p^*(\psi,\phi)q, \tag{1}$$

in conformity with $|\phi q\rangle=|\phi\rangle q$.

Under the initial assumptions of Wigner,[5] reformulated by Bargmann,[6] or the assumptions of Emch and Piron,[7] a symmetry μ is defined as a map that preserves transition probabilities between rays, or equivalently as an automorphism of the orthocomplemented lattice $\mathscr{P}(\mathscr{H}_{\mathbb{H}})$, the elements of which are the closed subspaces (i.e., the projectors) of the Hilbert space $\mathscr{H}_{\mathbb{H}}$.

The theorem known as Wigner's theorem (by physicists), and as the infinite-dimensional version of the fundamental theorem[8] of projective geometry (by mathematicians) asserts that every symmetry is implemented by a counitary operator U, satisfying

$$P\in\mathscr{P}(\mathscr{H}_{\mathbb{H}})\mapsto\mu[P]=U^*PU, \tag{2}$$

with

$$U(\psi q)=(U\psi)\alpha_U[q],\forall\psi\in\mathscr{H}_{\mathbb{H}} \quad\text{and}\quad q\in\mathscr{H}, \tag{3a}$$

and

$$\alpha_U[q]=\omega_U^*q\,\omega_U,\quad\text{for some } \omega_U\in\mathbb{H},\quad\text{with } \omega_U^*\omega_U=1; \tag{3b}$$

i.e., α_U is an automorphism of the field of quaternions. The counitarity of U means that

$$U^*U=UU^*=I,\quad\text{so that } (U\psi,U\phi)=\alpha_U[(\psi,\phi)], \tag{3c}$$

which reflects the fact that for a colinear operator A the adjoint is defined by

$$(A^*\psi,\phi)=\alpha_A^{-1}[(\psi,A\phi)]. \tag{4}$$

Conversely, every counitary operator implements a symmetry.

Finally, a symmetry determines the counitary operator that implements it, uniquely up to a "phase;" specifically the quaternionic form of Schur's lemma[1] implies that two counitary operators U_1 and U_2 implement the same symmetry if and only if there exists a unit quaternion ω, such that $U_2 = U_1 C_\omega$, where C_ω is the counitary operator defined by

$$C_\omega \psi = \psi \omega. \tag{5}$$

Indeed,

$$P \in \mathscr{P}(\mathscr{H}_{\mathrm{H}}) \mapsto C_\omega^* P C_\omega = P. \tag{6}$$

Hence, for every symmetry separately, one can choose a *unitary* operator to implement this symmetry; and this unitary operator is unique up to a sign.

So far, and as long as each symmetry is treated separately, the above approach covers the premises of both Adler[2] and Emch.[1]

II. STRONG AND WEAK PROJECTIVE REPRESENTATIONS

When an abstract group G is represented as a group of symmetries, i.e., when a symmetry $\mu(g)$ is assigned to every $g \in G$ in such a manner that

$$\mu(g_1)\mu(g_2) = \mu(g_1 g_2), \quad \forall (g_1, g_2) \in G \times G, \tag{7a}$$

i.e.,

$$P \in \mathscr{P}(\mathscr{H}_{\mathrm{H}}) \mapsto \mu(g_1)[\mu(g_2)[P]] = \mu(g_1 g_2)[P], \quad \forall (g_1, g_2) \in G \times G, \tag{7b}$$

one can repeat the above procedure for each g separately, and obtain a lifting by unitary operators $U(g)$, satisfying

$$U(g_1)U(g_2) = \pm U(g_1 g_2). \tag{8}$$

When G is a Lie group, and μ is a *continuous* representation, the brutal lifting just described may, however, not lead to a *continuous* unitary representation. As physics needs continuity to define the observables corresponding to the generators of the unitary representation, it is reassuring to know that continuity is obtained, nevertheless,[1] as a result of the following procedure.

First, one shows that there always exists a continuous local lifting by counitary operators, thus satisfying the condition

$$U(g_1)U(g_2) = U(g_1 g_2) C_{\omega(g_1, g_2)}. \tag{9}$$

In this expression C_ω is a counitary operator, defined as in (5), where now $\omega = \omega(\cdot,\cdot)$ is a continuous function of each of its arguments, takes its values in the unit quaternions, and satisfies, besides the trivial conditions $\omega(g,e) = \omega(e,g)$, the 2-cocycle condition:

$$\omega(g_1, g_2 g_3)\omega(g_2, g_3) = \omega(g_1 g_2, g_3)\alpha_{U_{g_3}^{-1}}[\omega(g_1, g_2)]; \tag{10a}$$

for the purpose of ulterior comparison with (13), we rewrite (10a) as

$$C_{\omega(g_1, g_2 g_3)} C_{\omega(g_2, g_3)} = C_{\omega(g_1 g_2, g_3)} U_{g_3}^{-1} C_{\omega(g_1, g_2)} U_{g_3}. \tag{10b}$$

Second, one shows that such a lifting is always equivalent to a *continuous, unitary, local*, but true *representation* (i.e., no ω, not even a $\pm$ sign, ambiguity).

Third, whenever the Lie group G is simply connected, this can be extended to a continuous, unitary representation of the whole group G. In cases where the group is doubly connected (e.g., the rotation group in three dimensions), one only obtains the above result for its covering group; it is when one has to consider the group itself that the $\pm$ ambiguity of (8) can possibly manifest itself. As the latter amendment (covering multiply connected groups) is not germane to the issue on which we want to concentrate in this paper, we will not pursue that part of the discussion here.

The straightforward generalization we just sketched, extending to quaternionic Hilbert spaces the analysis familiar from the complex Hilbert spaces situation presents one remarkable feature: the "phase reduction" is always locally trivial. Mathematically, this can be understood[1] from the fact that the local phase reduction amounts to finding, up to equivalence, all the extensions[9] of the Lie algebra of G by the Lie algebra of the group of automorphisms of the field of quaternions; as the latter happens to be the semisimple Lie algebra su$(2,C)$, all such extensions are trivial.[10] In this respect the complex case is much more involved, as shown by Bargmann.[11] In particular, the phase reduction is *not* locally trivial for the Galilei group, a fact that is interpreted as viewing the mass as parametrizing the sectors of a superselection rule. Two attitudes are possible in this juncture. The first, which was chosen by Emch,[1] was to accept that Galilean QM *is* different in its quaternionic realization from what it is in its complex realization. The second is to pursue the issue, and to generalize the definition of a projective representation; this was recently proposed by Adler.[2]

Translated in the notation of this paper, Adler's proposal[2] is to replace condition (9) by the weakened condition,

$$U(g_1)U(g_2)=U(g_1g_2)L_{\Omega(g_1,g_2)}, \tag{11}$$

where $L_{\Omega(g_1,g_2)}$ is the linear operator,

$$L_{\Omega(g_1,g_2)}\psi=\sum_k \phi_k\omega_k(g_1,g_2)(\phi_k,\psi), \tag{12}$$

with $\omega_k(g_1,g_2)^*\omega_k(g_1,g_2)=1$ and $\Phi=\{\phi_k|k=1,2,...\}$ is a complete orthonormal basis in $\mathscr{H}_{\mathrm{H}}$, the same for all pairs (g_1,g_2) of elements of G. Note that

$$L_{\Omega(g_1,g_2g_3)}L_{\Omega(g_2,g_3)}=L_{\Omega(g_1g_2,g_3)}U_{g_3}^{-1}L_{\Omega(g_1,g_2)}U_{g_3}. \tag{13}$$

III. DISCUSSION

While $\{11,13\}$ look somewhat similar to $\{9,10b\}$, there are major differences between these two formulations; our purpose in this paper is to delineate sharply the scope and reach of these variations.

First, (9) is a direct consequence of the condition (7). Hence one should expect condition (7) to be violated by (11). This is indeed the case: see (16) below. Recall that (7) is the defining condition for the usual definition of a projective representation, as $\mathscr{P}(\mathscr{H}_{\mathrm{H}})$ is the projective space associated to the vector space $\mathscr{H}_{\mathrm{H}}$. It is, in fact, equivalent to (9), and it is the condition Adler[2] refers to as the defining property of a *strong* projective representation, in opposition to (11), which is equivalent to (16), and which he introduces as the definition of a *weak* projective representation.

Second, (9) is a relation among essentially counitary operators. It is true, as we just mentioned, that a powerful theorem[1] allows us to reduce the phases and thus to obtain a locally trivial continuous unitary representation, so that (9) becomes ultimately a relation between linear operators. Nevertheless, this reduction is not instructive in the present juncture since it is (9) itself [not (8)] that serves as a motivation for the extension (11). By contrast, (11) is in its very essence a relation between unitary operators; in particular, L is a linear operator (in fact, a unitary operator)

that involves the choice of a complete orthonormal basis $\Phi=\{\phi_k|k=1,2,...\}$; i.e., the focusing on one complete set of commuting observables, or more precisely, on a discrete, maximal Abelian, real subalgebra,

$$\mathscr{A}_\Phi=\left\{A:\psi\in\mathscr{H}_{\mathrm{H}}\mapsto A\psi=\sum_k \phi_k a_k(\phi_k,\psi)\in\mathscr{H}_{\mathrm{H}}\right\}, \tag{14}$$

the minimal projectors of which are the projectors P_{ϕ_k} on the one-dimensional rays corresponding to each element ϕ_k of the chosen basis Φ. We denote by $\mathscr{P}(\mathscr{A}_\Phi)$ the Boolean sublattice of $\mathscr{P}(\mathscr{H}_{\mathrm{H}})$ generated by these projectors.

Third, as a consequence of the above remark, whereas the colinear operators $C_{\omega(g_1,g_2)}$ in (9) implement the trivial symmetry [see (6)]—and are, in particular, independent of any choice of a Hilbert space basis—that is not the case for the symmetry implemented by the linear operators $L_{\Omega(g_1,g_2)}$. Indeed, we have generically only

$$P\in\mathscr{P}(\mathscr{A}_\Phi)\mapsto L^*_{\Omega(g_1,g_2)}PL_{\Omega(g_1,g_2)}=P. \tag{15}$$

Hence, the symmetry implemented by $U(g_1g_2)$ coincides with the symmetry implemented by $U(g_1)U(g_2)$ only for the elements of the distinguished maximal Abelian algebra $\mathscr{A}_\Phi$ chosen to define the linear operators $L_{\Omega(g_1,g_2)}$:

$$P\in\mathscr{P}(\mathscr{A}_\Phi)\mapsto\mu(g_1)[\mu(g_2)[P]]=\mu(g_2g_2)[P],\quad \forall(g_1,g_2)\in G\times G. \tag{16}$$

This, compared to (7), is the major difference between the conditions defining weak versus strong projective representations. While both require, for each symmetry *separately*, that $\mu(g)$ be an automorphism of the *whole* system (a condition necessary to support the use of Wigner's theorem), the difference appears when it comes to the representation of a *group* of symmetries: the strong definition requires (7b), i.e., that μ is a representation on the full $\mathscr{P}(\mathscr{H}_{\mathrm{H}})$, whereas the weak definition requires only (16), i.e., that this condition hold on $\mathscr{P}(\mathscr{A}_\Phi)$.

This is the price one must be prepared to pay for the relaxing from the ''strong'' condition (9) to the ''weak'' condition (11)—which is the generalization proposed by Adler.[2] At this price, it has become possible[12,4,13] to classify the irreducible weakly projective representations of connected Lie groups; to embed complex projective representations into weakly projective quaternionic representations (even when the Bargmann complex phase reduction is not locally trivial); to construct quaternionic coherent states (including the weakly projective case); and to discuss how, in the complex case, the weak condition (11) already implies the stronger condition of (9).

After comparing their original motivations, the authors realized how they both had hoped to take advantage of the SU(2) symmetry of the quaternions: Emch[1] was interested in finding some natural coupling between the inhomogeneous Lorentz group of special relativity and the internal symmetries then known in elementary particle theory; Adler[2] was similarly interested in finding a source in the ray structure of Hilbert space for the color symmetry. It seems fair to say that, even with the generalization proposed by Adler,[2] the structure of the current quaternionic models for quantum theories is not (yet) rich enough to accommodate dreams that extend beyond the complex Hilbert space formalism.

ACKNOWLEDGMENTS

The authors thank Dr. A. Jadczyk for discussions on matters related to this paper.

The work of S. L. Adler was supported in part by the Department of Energy under Grant No. DE-FG02-90ER40542.

[1] G. G. Emch, "Mécanique quantique quaternionienne et relativité restreinte. I." Helv. Phys. Acta **36**, 739–769 (1963); **36**, 770–788 (1963); "Representations of the Lorentz group in quaternionic quantum mechanics," presented at the *Lorentz Group Symposium* Summer 1964, Lectures in Theoretical Physics, edited by W. E. Brittin (University of Colorado Press, Boulder, CO, 1964), Vol. VII, pp. 1–36.

[2] S. L. Adler, "Projective group representations," J. Math. Phys. **37**, 2352–2360 (1996); *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford University Press, New York, 1995).

[3] G. G. Emch, "Comments on a recent paper by S. Adler on projective group representations in quaternionic Hilbert spaces," J. Math. Phys. **37**, 6582–6585 (1996).

[4] S. L. Adler, "Response to the Comment by G. Emch on projective group representations in quaternionic Hilbert space," J. Math. Phys. **37**, 6586–6589 (1996).

[5] E. P. Wigner, *Gruppentheorie* (Vieweg und Sohn, Braunschweig, Germany, 1931); *Group Theory* (Academic, New York, 1959).

[6] V. Bargmann, "Note on Wigner's theorem on symmetry operations," J. Math. Phys. **5**, 862–868 (1964).

[7] G. Emch and C. Piron, "Note sur les symétries en théorie quantique," Helv. Phys. Acta **35**, 542–543 (1962); "Symmetry in quantum theory," J. Math. Phys. 4, 469–473 (1963).

[8] E. Artin, *Geometric Algebra* (Interscience, New York, 1957).

[9] H. Cartan and S. Eilenberg, *Homological Algebra* (Princeton University Press, Princeton, NJ, 1956).

[10] N. Bourbaki, *Eléments de Mathématiques, Groupes et Algèbres de Lie, Fasc. XXVI* (Hermann, Paris, 1960).

[11] V. Bargmann, "On unitary ray representations of continuous groups," Ann. Math. **59**, 1–46 (1954).

[12] T. Tao and A. C. Millard, "On the structure of projective group representations in quaternionic Hilbert space," J. Math. Phys. **37**, 5848–5857 (1996).

[13] S. L. Adler and A. C. Millard, "Coherent states in quaternionic quantum mechanics," J. Math. Phys. **38**, 2117–2126 (1997).

Stephen L. Adler: Vita

Address

Institute for Advanced Study
Einstein Drive
Princeton, NJ 08540, USA
adler@ias.edu

Personal Information

Born in New York City, November 30, 1939
A.B. Harvard University, 1961
Ph.D. Princeton University, 1964

Affiliations

Junior Fellow, Society of Fellows, Harvard University, 1964–66
Postdoc, California Institute of Technology, 1966
Institute for Advanced Study:
  Member, 1966–69;
  Professor of Theoretical Physics, 1969–
Fellow, American Academy of Arts and Sciences
Fellow, American Association for the Advancement of Science
Fellow, American Physical Society
Member, National Academy of Sciences

Awards

J.J. Sakurai Prize, American Physical Society, 1988
Dirac Medal of The Abdus Salam International Centre for Theoretical Physics, 1998

Home page

For a more detailed vita, and a full bibliography, see:
http://www.sns.ias.edu/~adler/